WIDESPREAD BLOCKS OF GENOMIC HYPOMETHYLATION IN HUMAN AGING AND SUN EXPOSURE

by

Amy Ruth Vandiver

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

May, 2015

Abstract

Human aging is the greatest risk factor for disease and mortality, however the molecular mechanisms are poorly understood. The process involves altered cellular function but occurs without predictable genetic change and is variable among individuals, indicating that epigenetic change may be involved. DNA methylation is an epigenetic mark with important roles in regulating gene expression and cellular identity. Widespread differences in DNA methylation have been observed in cancer, EBV transformation of lymphocytes and replicative senescence. Early work considering epigenetic change in aging identified global changes in methylation highly variable among individuals. Recent work has identified specific genomic regions in which DNA methylation consistently changes with age, however the genome scale patterns and the factors underlying this change remain unclear.

In this work, we use genome wide methods to clarify the nature of age related epigenetic change. Using human skin samples, we identify widespread genomic blocks of hypomethylation in chronically sun-exposed epidermis but not sun-protected, aged tissue, implicating environmental stress in mediating large-scale epigenetic change. The degree of hypomethylation in the identified blocks correlates with clinical measures of photo-aging and the identified blocks are further hypomethylated in squamous cell carcinoma samples compared to matched controls.

In another study, using purified monocytes and lymphocytes from human peripheral blood, we observe highly cell type specific patterns of methylation change with age. In lymphocytes, we observe widespread blocks of hypomethylation similar to

the changes identified in cancer and chronically sun-exposed epidermis, while in monocytes, no large-scale change is present with age. In addition, we identify smaller regions differentially methylated with age in each cell type that occur in cell type specific enhancers and contain motifs for regulatory factor binding, indicating methylation change with age may be linked to the regulation of cell identity.

Given the widespread blocks of hypomethylation observed in skin and blood systems, we investigate the link between block methylation and proliferation by using whole genome bisulfite sequencing to measure DNA methylation in growth arrested and actively dividing primary fibroblasts. We demonstrate that methylation in block regions is stable over replication and through extended arrest.

Readers:

Winston Timp, Ph.D.

Andrew P. Feinberg, M.D., M.P.H.

Thesis Advisor:

Andrew P. Feinberg, M.D., M.P.H.

Acknowledgements

First and foremost I would like to thank my advisor, Andrew P. Feinberg, for giving me the opportunity to work in his laboratory and mentoring me. From the moment I met him, he has believed in me, encouraged me and challenged me. His creativity and true love of science are truly inspiring and infectious. I will never forget everything he has taught me.

I am very grateful to my committee members, Luis Garza, Sean Taverna, Winston Timp and Jeremy Walston for all their guidance and support. They have given me invaluable feedback, direction and encouragement. It has been an honor to work with such insightful and accomplished advisors.

In addition to my committee members, Rafael Irizarry and Kasper Hansen have provided invaluable guidance in biostatistics, Anna Chien and Sewong Kang have taught me a great deal about dermatology, Robert Siliciano and Andrea Cox have been very helpful in planning my career.

Working among the members of the Feinberg lab and the Epigenetics Center has been one of the best experiences of my life. I have been able to go to work every day and interact with brilliant, kind and fun individuals. I am especially grateful to my fellow Feinberg lab graduate students: Namyoung Jung, Brian Herb, Carolina Montano, Akiko Doi, Michael Multhaup, Elisabet Pujadas, Varenka Rodriguez and Hwajin Lee for teaching me, developing protocols with me, giving endless feedback and always being available to encourage and comfort me when work was difficult. Our postdocs, Yun Liu,

Xin Li, Tal Salz, Lindsay Rizzardi and Kelly Bakulski, have been amazing resources for me as well.

I would not have made it through my PhD work without the support of my friends. At the beginning of graduate school, I began volunteering with Back on My Feet Baltimore. I have met so many wonderful friends through this organization that have made Baltimore feel like home. Abbie Day, Sadie Smith, Caroline Speirs, Meredith Costa, Erin Leech and others have brought a lot of fun to my time as a PhD student.

Above all, my accomplishments are a testament to my amazing family. My parents, Tom and Janet Unterman, and my older sister, Rebecca Unterman, have always believed in me and supported me when I got the crazy idea of pursuing a PhD in the middle of my medical education. They are three of the most intelligent, considerate and brave people I have encountered. I could not ask for better role models. My husband, Scott Vandiver is the best partner I could imagine. He is brilliant, level-headed and fun. Having him to go home to has given me courage to challenge myself and makes every day worthwhile.

Chapter 1 contains material previously printed in Genome Biology (Vandiver et al. 2015).

Unpublished data contained within this work is property of the Feinberg lab.

Table of Contents

List of Tables

List of Figures

Introduction

DNA Methylation

Epigenetics is the study of heritable regulation of the transcriptional program that is independent of the DNA sequence itself. Conrad Waddington introduced the concept in the context of developmental biology. He noted that in a multicellular organism, a single cell develops into many cell types with distinct functions and identities, despite sharing an identical genotype. He proposed another level of information, the "epigenotype" as the mediator of this relationship between genotype and phenotype [1]. Today, epigenetic regulation is studied on multiple levels: DNA modifications to the DNA itself, modifications to the histone proteins around which DNA is wrapped, inter- and intra-chromosomal interactions, and localization of regions of DNA within the nucleus.

DNA methylation is the most commonly studied epigenetic modification. DNA methylation refers to the addition of a methyl group to the 5' position of the cytosine base. In mammals, it is most commonly found in the context of CpG dinucleotides, although cytosine methylation in other contexts has recently been reported in brain tissue and embryonic stem cells [2, 3]. CpGs are relatively depleted across vertebrate genomes, but are enriched in regions termed "CpG Islands", which have a >50% CG content, average 1000 bp and are most commonly found in gene promoters [4, 5]. While CpGs across the genome are generally heavily methylated, CpGs within islands have low levels of methylation, leading to a bimodal distribution pattern of methylation levels.

The methylation mark is placed on DNA by methyltransferase enzymes that catalyze transfer of the methyl group from S-adenosyl methionine. In humans, three methyltransferase enzymes have been identified. DNMT1 functions as the maintenance

methyltransferase; it catalyzes methylation of hemimethylated DNA to maintain the existing pattern of methylation during cell division. This enzyme localizes to newly synthesized DNA through association with the replication complex during DNA replication; it also associates with other transcription factors outside of replication [6, 7]. The *de novo* methyltransferases, DNMT3a and DNMT3b, function in *de novo* methylation. These enzymes have both overlapping and distinct roles: both enzymes function in methylation of repeat elements and in CpG island methylation, while DNMT3b is specifically implicated in methylation at pericentromeric repeats and in intergenic methylation [8-10].

Methylation at a given location can be lost through multiple pathways. Active demethylation occurs during development through hydroxylation by the Tet enzymes [11]. In addition, there is evidence of demethylation via deamination by AID followed by basis excision repair [12]. DNA methylation can be lost passively during DNA-replication if DNMT1 fails to properly localize [13]. In development, this passive loss may be partially dependent on initial active demethylation by Tet [14, 15].

The presence of DNA methylation at CpG islands in gene promoters is linked to decreased expression of nearest genes. The presence of methylation can directly inhibit binding of some transcription factors [16, 17]. DNA methylation can also act through recruitment of the methyl dependent binding protein MeCP2, which recruits chromatin modifiers to repress transcription by condensing chromatin [18, 19]. By contrast, methylation of intergenic CpG islands is associated with active transcription [20]. Methylation found outside of genes is thought to function both in repression of

transcription of repeat sequences at centromeres, telomeres and repetitive elements and in control of recombination [21-24].

In recent years, rapidly developing array and sequencing technology has facilitated characterization of genome-wide methylation in a large variety of samples, and has thus provided insight into where methylation differs between cells and how methylation changes with disease. While previous methods were restricted to measuring methylation at CpG islands or repeat elements, the development of the comprehensive high-throughput arrays for relative methylation (CHARM, detailed below), technology in 2009 allowed an unbiased comparison of genome-wide methylation between multiple tissue types. Using this method, Irizarry et al. observed that the regions differentially methylated between tissues are most frequently found not in CpG islands, but in the regions neighboring islands, termed "CpG shores" [25]. The differentially methylated shores identified in this work showed a strong inverse relationship between methylation and gene expression, indicating that these shore regions are an important location for regulation of gene expression. Strikingly, comparing colon cancer and normal samples on the same array revealed cancer related differential methylation occurs in similar locations to those that differentiate tissues[25]. Further, comparing induced pluripotent stem cells and embryonic stem cells to fibroblasts revealed differentially methylated regions in the same CpG shores, suggesting that methylation at CpG shores plays a role in determining cellular identity and function [26].

The development of whole genome bisulfite sequencing (WGBS, detailed below) technology has allowed genome scale observations of where differential methylation occurs. Through WGBS, large scale patterns of change have been observed in addition to

the small differentially methylated regions identified with array technology. Sequencing of cultured fibroblasts revealed that in addition to the expected distribution of highly methylated and primarily unmethylated regions, large domains with intermediate levels (~70%) of methylation are present, termed partially methylated domains (PMDs) [27]. When colon cancer and normal colon samples were compared, widespread hypomethylation was observed in colon cancer, however this change was not uniform across the genome. The majority of hypomethylation in colon cancer samples was found to occur in large regions with low CpG and low gene density, called hypomethylated blocks [28]. These domains were observed to overlap strongly with the PMDs from fibroblasts. Further, they overlap with domains previously identified to associate with the nuclear membrane (LADs) and regions of condensed chromatin (LOCKs), both of which are associated with decreased transcription and have been implicated in regulation of cellular identity [29, 30]. A later study identified similar hypomethylated blocks associated with EBV transformation of lymphocytes, indicating that loss of methylation in blocks may occur early in the process of oncogenesis [31].

In addition to its roles in embryonic development and cancer, maintenance of the methylome is essential for maintaining regenerative potential in adult progenitor cell populations and *de novo* methylation plays an important role in differentiation. A comprehensive mapping of whole genome methylation in hematopoietic stem, progenitor and daughter cell populations demonstrated a highly specific process of de-methylation and re-methylation during differentiation [32]. The use of conditional knockouts has demonstrated that loss of DNMT1 leads to loss of regenerative capacity in hematopoietic and keratinocyte stem cells and to loss of differentiation capacity in intestinal crypt stem

cells [33-35]. The *de novo* methyltransferases are implicated in mediating differentiation of regenerating cells: loss of DNMT3a and 3b in hematopoietic stem cells impairs differentiation [36], and DNMT3a is essential for osteoclast differentiation from precursor cells [37].

The mechanism of region-specific differential methylation is not understood. One study indicates binding of transcription factors at regulatory regions leads to local hypomethylation, suggesting a cell's methylome may reflect which regulatory factors have recently been DNA bound [38]. Other work indicates that the *de novo* methyltransferases are recruited in complex with enzymes that promote heterochromatin formation [39, 40]. Given the strong overlap between hypomethylated blocks, lamina-associated domains and heterochromatin domains, it is possible that altered methylation in these regions is a result of large scale nuclear reorganization.

Methods to study genome-wide methylation

The CHARM array, developed in 2009, offered one of the first opportunities to look at methylation genome-wide in an unbiased manner. In this method, genomic DNA is sheared and half of the DNA is digested with McrBC, a restriction enzyme that cuts only when bound to methylated DNA. After size selection, the digested and undigested DNA is then purified, amplified and hybridized to a tiling array, covering the genome without bias towards CpG islands. The relative methylation at a given region is determined by comparing hybridization signal for each region from the MCRBC digested DNA to the undigested DNA: a weaker signal indicates more digestion and thus more methylation [41].

The Illumina HumanMethylation27 beadchip (27k) and HumanMethylation450 beadchip (450k) allow quantitative measurement of methylation at specific CpGs by applying genotyping technology to DNA treated with sodium bisulfite. The bisulfite treatment deaminates unmethylated cytosines to uracil. When this DNA is PCR amplified, methylated cytosines are amplified as cytosine, while converted, unmethylated, cytosines are amplified as thymidine. The Illumina arrays then measure the percentage of cytosines compared to thymidines at specific CpGs using preferential hybridization and extension [42]. The original 27k array measured only cytosines within CpG Islands, limiting its application for genome wide studies. The 450k array was developed after the CHARM array and was designed to assay CpGs within CpG islands, shores, and other regions shown to variably methylated across tissues, cell types and populations [43].

The development of whole genome bisulfite sequencing (WGBS) technology has allowed genome scale observations of where differential methylation occurs. In this method bisulfite converted, PCR amplified DNA is sequenced. This allows quantitation of methylation at every cytosine across the genome by comparing the number of cytosine and thymidine reads at each location [44]. While sequencing for genotype requires relatively high coverage, the BSmooth algorithm allows accurate quantitation of methylation at base-pair resolution from lower coverage WGBS data by borrowing information from neighboring CpGs [45], making WGBS an accessible option for analyzing a many kinds of samples.

Aging

In humans, aging is the greatest risk factor for disease and morbidity, affecting all organ systems in similar and tissue specific ways [46]. Within organ systems, the changes

of aging are often explained as a loss of homeostasis arising from altered cellular function. In many settings, the cellular changes involve loss of function: in muscles, hair follicles, bone and cardiac tissue, cells lose their ability to regenerate tissue [47-50]; in the dermis, fibroblasts decrease production of collagen [51]. However, many aging phenotypes are also attributed to increased cell activity: there is an increase in cancer and benign proliferations and inflammatory cells exhibit higher basal levels of activation [52]. These changes are widely attributed to the accumulation of cellular damage over time, but the nature of damage and the mechanism of this accumulation remain unclear.

Both loss and gain of cellular function with aging may be linked to the accumulation of senescent cells in tissue. Cellular senescence, a protective response in which the cell exits the cell cycle, can be induced in culture by serially passaging of primary cells or as an acute stress response (reviewed in [53]). Senescent cells can be visualized in tissue by β-gal staining or p16 expression and are observed to accumulate with age in human skin, kidney, cardiac, muscle and other tissues [47, 54-56]. Senescent progenitor cells fail to properly regenerate tissue, leading to many loss of function phenotypes [47]. Further, the presence of senescent cells in tissue stimulates local inflammation and tissue remodeling, which may lead to many of the hyper-proliferative and inflammatory phenotypes associated with age [57-59].

In addition to the presence of senescent cells, alterations to chromatin structure and other cellular changes contribute to aging phenotypes across many systems. In 2013, Otin et al. proposed nine “hallmarks of aging” which demonstrably contribute to lifespan. These include genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence,

stem cell exhaustion and altered intercellular communication [60]. With regard to epigenetic change, decreased global levels and altered distribution of heterochromatic marks are associated with aging in c. elegans and drosophila [61, 62]. Functional relevance of these changes is supported by studies in which deletion of histone K27 or K4 methylation machinery or manipulation of heterochromatin binding proteins directly alters lifespan [62-65]. Otin et al. note that, in addition to chromatin changes, consistent DNA methylation changes have been observed with mammalian aging, however these changes have not yet been directly implicated in regulation of lifespan [60]. These observations are discussed in detail below.

In this work, we focus on two tissues allowing examination of different aspects of human aging: human skin, in which the majority of aging phenotypes are driven by UV and other environmental exposures, and human peripheral blood, where aging phenotypes are primarily attributed to changes in function of the progenitor cell population. Aging in both tissues has been extensively studied, below is a brief introduction to the changes that have been observed in each setting.

Aging in human skin

Mammalian skin consists of two layers. The superficial layer, the epidermis, is approximately 60 μm thick and consists of stratified layers of keratinocytes, regenerated by the basal layer. In addition, for approximately every 36 keratinocytes, there is 1 melanocyte and for every 53 keratinocytes, there is 1 Langerhans cell [66]. Melanocytes are found within the basal layer and produce melanin pigment, which spreads through the epidermis and acts as both a physical barrier to UV penetration of the skin and as a scavenger of reactive oxidative species [67, 68]. The Langerhans cells are immature

dendritic cells that internalize foreign proteins and act as antigen presenting cells. The deeper dermal layer consists of a dense matrix of connective tissue, primarily collagen and elastin. The most prevalent resident cell population in the dermis is dermal fibroblasts that produce this matrix.

Skin aging is intimately entwined with the effects of chronic UV exposure. The histological changes associated with aging and sun exposure in skin have been extensively studied. The primary histopathological changes associated with aging and sun exposure are structural: In the epidermal layer, increased thickness is associated with sun exposure; decreased thickness is associated with aging. Within the dermal layer, loss of collagen and elastin and altered fibroblast morphology is associated with both chronological aging and sun exposure [69, 70]. Both extrinsic and intrinsic skin aging are associated with broad changes in gene expression, with many similar pathways differentially regulated in each. For example, loss of dermal collagen in both sun-protected and sun-exposed skin is associated with increased expression of matrix metalloproteinases in the epidermis downstream of AP-1 signaling. Nuclear localization of AP-1 is activated in sun-exposed skin by UV exposure and is also increased with age in sun-protected skin [71].

In addition to structural changes, there is evidence of altered function of skin with age. Aged skin has decreased barrier function, as evidenced by water loss and permeability, and delayed regeneration after wounding [72]. This change is attributed to alterations within the extracellular matrix of the dermis, decreased regenerative capacity of keratinocytes and altered lipid profiles within the epidermis [72-74]. In parallel to this decreased function, there is an increasing risk of skin cancer with age, especially basal

cell and squamous cell carcinoma, which arise from the keratinocyte cells of the epidermis [75].

Aging in human blood

Peripheral blood consists of multiple cell types regenerated from hematopoietic stem cells (HSCs) residing primarily in the bone marrow. The largest population of cells is enucleate red blood cells, which function in carrying oxygen to tissues and transporting carbon dioxide away from tissues. Within the nucleated cell fraction, the most common population is polymorphonuclear neutrophils (~60%), which function in the innate immune response. The next most prevalent are lymphocytes (~30%), which function in the adaptive immune response. Within the lymphocyte population, there are naïve cell populations, which respond to new antigens, and memory cells that mediate quick responses to previously encountered antigens. Approximately 5% of cells in the peripheral blood are monocytes, which migrate into target tissues and function in the innate response in those tissues. Other cell types within the peripheral blood include basophils, which release histamine in certain nonspecific immune reactions and eosinophils, which respond to parasitic infection.

Aging in the hematopoietic system is associated with decreased ability to respond to new infections, termed immunosenesence [76]. This is linked both to altered differentiation capacity of hematopoietic stem cells and altered function of differentiated cells. As mammals age, the HSC population increases, however it does not continue to replenish all cell types evenly. There is decreased production of lymphocytes and increased production of myeloid cells [77, 78]. In addition, lymphocytes present within the elderly have a decreased ability to coordinate response to new infection, due to

decreases in the number of naïve cells and impaired interactions between mature lymphocytes [79].

In addition to immunosenesence, hematopoietic aging is associated with increased basal activation of inflammatory pathways. Older subjects have repeatedly been observed to have elevated pro-inflammatory cytokines in serum, the levels of which correlate with decreased functional capacity with age [80]. *In vivo* work indicates this is due in part to cell intrinsic changes, as monocytes from aged individuals show increased inflammatory cytokine production after stimulation [52]

DNA Methylation change with aging:

Both regional and large-scale DNA methylation changes associated with aging have been reported in multiple tissues. Before the development of array technology, multiple groups reported decreased methylation in repeat elements associated with aging of peripheral blood [81, 82]. A study of two family cohorts sampled at multiple time points demonstrated change to global methylation levels with age, however the direction and magnitude of change was seen to cluster among families, implicating genetic influence on global methylation over time [83]. By contrast, an early study of global and site specific methylation of monozygotic twins showed a divergence in methylation over lifetimes, indicating environmental influence [84].

The development of array and sequencing technology has allowed more nuanced analysis of where in the genome DNA methylation changes with age. An early study using the Illumina HumanMethylation27 Beadchip (27k) array, a promoter-biased array, identified common sites of differential methylation with age in CD4+ lymphocytes and CD14+ monocytes. In this study, age associated hypermethylation occurred preferentially

in regions with both repressive and permissive chromatin marks ("bivalent domains") in embryonic stem cells and regions with repressive chromatin marks in differentiated cells [85]. The release of the more genome scale Illumina HumanMethylation Beadchip450 (450k) array has facilitated more thorough observations. 450k array studies in peripheral blood, brain and muscle have confirmed the association between hypermethylated age-DMRs and regions found to have repressive chromatin marks (H3K27 trimethylation and H3K9 trimethylation), and have observed that regions hypomethylated with age are frequently marked by H3K4 monomethylation [86-88].

Utilizing the 450k methylation data available from multiple tissues, Horvath identified a model using 353 CpGs that is used to calculate a "methylation age". This age correlates strongly with chronological age across a large variety of tissues and cell types[89]. The "methylation age" determined by applying this model to peripheral blood data was observed to predict with mortality and functional measures [90, 91] and accelerates in some systems of accelerated aging [92, 93]. This work suggests that aging of the peripheral blood methylome may serve as a marker of biological aging.

The use of whole genome bisulfite sequencing to study primary tissue across ages has been limited, but indicates that more large scale change maybe present. Sorted CD4+ cells from one newborn and one centenarian examined by Heyn et al. using whole genome bisulfite sequencing (WGBS) showed large scale hypomethylation with age, primarily in repeat-rich and gene-poor regions, however the generalizability of this result is limited by the use of only one cell type and single donors in each age group [94]. Intriguingly, widespread blocks of hypomethylation were also observed *in vivo* in fibroblasts cultures to replicative senescence [95]. These findings suggest the early

observations of hypomethylated repeat elements with age may be linked to the development of hypomethylated regions similar to the blocks observed in cancer.

Human skin offers a particularly attractive model for studying the environmental contribution to methylation change with aging, however studies of DNA methylation in skin aging have been limited. An early study used the 27k array and saw few changes associated with either aging or exposure, possibly due to the limited coverage of this array [96]. A more recent study used Whole Genome Bisulfite Sequencing (WGBS) to examine methylation in aging skin more comprehensively, however they looked only at samples from one body site, the inner forearm, of each individual and thus were not able to characterize the interaction between intrinsic and extrinsic aging in these samples[97].

While changes to the DNA methylome with age have been observed across many systems, the underlying mechanism of this change has not been established. Given the overlap between age- hypermethylated cytosines and genomic regions marked by H3K27 and H3K4 trimethylation, it is possible that the age related loss of enzymes placing those marks, as observed in flies and c. elegans, leads to DNA methylation changes by altering local chromatin structure and/or transcription factor binding. Loss of methylation at the large block domains was attributed to altered DNMT1 localization in late passage, senescent fibroblasts [95], however the cause of this change and relevance to *in vivo* aging is unknown.

The functional consequences of the commonly reported age related methylation changes are similarly unknown. While differential methylation between tissues correlates with differential gene expression, there is no evidence of this relationship for CpGs identified as determining “methylation age” [89]. A study comparing methylation

and gene expression of purified monocytes and lymphocytes from older donors further demonstrated that regions differentially methylated with age across multiple cell types do not show a clear relationship to expression of the nearest gene [98]. In a study of hematopoietic stem cell aging, no relationship was seen between regions differentially methylated and genes differentially expressed with age in mice. However, that study noted that the methylation changes occurred near genes required for downstream differentiation, suggesting that differential methylation stem cells may regulate their ability to induce appropriate gene expression during differentiation or in response to stimuli [99].

In this work, we use unbiased, genome-wide methods to characterize DNA methylation change with age and investigate the factors underlying widespread change. By using the 450k array and WGBS to analyze human epidermal and dermal samples from older and younger individuals, we examine the influence of age and sun-exposure on the methylome and implicate chronic exposure in widespread epigenomic change. To clarify the cell intrinsic nature of epigenomic change in a non-exposure driven model, we perform a multi-scale analysis of DNA methylation in sorted blood cell populations from younger and older donors, demonstrating cell type specificity in both global and local methylation change with age. Given the widespread changes in methylation observed in these systems, we examine growth arrested and dividing cells in culture to determine whether cell proliferation may underlie our findings.

Chapter 1: Age and sun-exposure related genomic blocks of hypomethylation in nonmalignant epidermis

The material presented in this chapter has been published in Genome Biology (Vandiver et al. 2015) and is reproduced in part below.

Introduction

Aging is the greatest single risk factor for cancer, cognitive decline, frailty, and immunological dysfunction [46], yet consistent genomic alterations related to aging have been elusive. For example, few genetic variants regulating human life span have been identified [100, 101]. At the same time, there is a growing realization that environmental factors are major contributors to aging and age-associated illness. Epigenetics is the study of chemical modifications of the genome, heritable by cell progeny, and it has been an attractive target for studies of aging and environmentally-influenced disease. Several groups have shown differences in DNA methylation—a covalent modification of cytosine at CpG dinucleotides—in peripheral blood samples and other tissues with increasing age [85, 94, 102]. Some of these differences are possibly confounded by changes in cell type distribution with aging [103], but many are likely real as they are seen across multiple cell types. Studies of identical twins have shown markedly divergent patterns of DNA methylation in whole blood over the lifespan, suggesting an environmental component to epigenetic change with age [84], and the epigenetic drift hypothesis [104].

The skin as a model of aging offers the advantage of studying the influence of environmental factors by virtue of its direct exposure to the sun. The superficial layer of epidermis (~60 μm) interfaces more directly with the outside world than the deeper dermal layer. Even penetration of solar ultraviolet (UV) radiation is mostly in the epidermis [105]. Furthermore, skin affords the ability to compare the effects of intrinsic and extrinsic (environmental) aging through the comparison of chronically sun-exposed

(e.g., forearm and face) and sun-protected skin (e.g., upper inner arm) in the same individual. Both layers offer a relatively homogenous system for analyzing DNA methylation; the epidermis especially is composed of around 95% keratinocytes [106]. Histological changes associated with aging and sun exposure in human skin have been extensively studied. The primary histopathological changes associated with aging and sun exposure are independent of cell type change: In some studies of the epidermal layer, increased thickness is associated with sun exposure, decreased thickness is associated with aging. Within the dermal layer, loss of collagen and altered fibroblast morphology is associated with both chronological aging and sun exposure [69, 70]. Despite the lack of large cell type shifts, both extrinsic and intrinsic skin aging are associated with broad changes in gene expression, with many similar pathways differentially regulated in each. Pathways associated with intrinsic aging in sun-protected skin are seen to be amplified by chronic environmental exposure in sun-exposed skin [71, 107]. We therefore hypothesized that epigenetic changes might mediate the changes associated with environmental exposure in aging skin.

A previous study of DNA methylation in aging skin tissue was limited technologically. This study used the Infinium HumanMethylation27 BeadChip which measures 27,000 CpG sites, focused on dense CpG regions termed CpG islands [96]. It showed little change related to skin aging (hypermethylation at 0.38% of sites) and even less change associated with sun exposure (hypomethylation at 0.05% of sites)[96]. We and others have recently observed widespread differentially methylated regions (DMRs) across the genome that distinguish tissues (t-DMRs), stages of stem cell reprogramming (r-DMRs), and cancer (c-DMRs). Most of these alterations are either at regions near but

not in CpG-dense islands, termed CpG island shores, or distal from both islands and shores (the "open seas") [25, 26, 28, 32]. These distal regions were discovered to be large blocks, corresponding to heterochromatin regions termed large organized chromatin lysine-modifications (LOCKs) or nuclear lamin-associated domains (LADs), and these large blocks show substantial hypomethylation in cancer [28, 108]. Almost none of these regions is represented on the Infinium HumanMethylation27 BeadChip, and thus were not included in the previous study. A more recent study used Whole Genome Bisulfite Sequencing (WGBS) to examine methylation in aging skin more comprehensively, however they looked only at samples from one body site, the inner forearm, of each individual and thus were not able to characterize the interaction between intrinsic and extrinsic aging in these samples[97].

In order to more fully examine methylation in human skin samples affected by age and sun-exposure, we performed array-based DNA methylation analysis using the Infinium HumanMethylation450 BeadChip array, which includes islands, shores, most known c-DMRs, t-DMRs, and r-DMRs and probes within the "open sea" regions identified as part of the hypomethylated blocks in cancer [109]. We applied recently developed algorithms [110] capable of identifying differences in large blocks as have previously been detected in cancer samples using whole genome bisulfite sequencing (WGBS), and we confirmed these results directly with WGBS of additional samples. Here we describe profound changes in DNA methylation associated with combined aging and sun exposure, involving 670 Mb of the genome, and including large blocks similar to cancer. In human subjects, these changes are progressive with quantitative measures of

skin aging. These same blocks are observed to be hypomethylated in squamous cell carcinoma samples.

Results

We undertook a comprehensive genome-wide analysis of DNA methylation in human skin to test specifically for alterations associated with age or with sun exposure, and the potential relationship between these regions. Ten younger individuals (<35 years old) and 10 older individuals (>60 years old) each submitted one sun-protected biopsy specimen from the upper inner arm, as well as one sun-exposed specimen from either the dorsal forearm or crow's feet (lateral epicanthus) (donor information in Table 1.1). Donors were carefully selected to exclude individuals with active skin conditions or individuals using topical medications (exclusion criteria detailed in methods). As methylation and sun-exposure effects both vary with race, we limited this sample set to Caucasian donors. In addition to tissue biopsies, donors volunteered health history information. For each donor, a dermatologist evaluated their degree of apparent skin aging in both sun-exposed and sun-protected regions using two established scales.

In order to study more homogenous tissue samples, each punch biopsy was mechanically separated into epidermis and dermis following overnight incubation with dispase, a procedure associated with unappreciable fibroblast contamination in cultured epidermal sections [111], and analyzed separately. In previous work, epidermal sections separated in this manner show the same age and sun exposure related methylation changes as epidermal sections separated by suction blister, indicating the dispase separation does not have significant effects on the methylome [96]. To confirm uniformity of our technique among sample groups, we sectioned a subset of dispase

separated epidermal layers and performed hematoxylin and eosin staining. Examination of stained sections showed separation below the basal layer for all age and sun exposure groups, consistent with previous histological analysis of dispase separated epidermis (Figure 1.1) [112]. To characterize methylation genome-wide, we used the Infinium HumanMethylation450 bead chip, which includes most CpG islands and shores, additional regions shown to be differentially methylated in cancer and development, and other functionally important regions [109].

We classified the samples from each skin layer into four groups: younger sun-protected (Y-pro), younger sun-exposed (Y-exp), older sun-protected (O-pro), and older sun-exposed (O-exp). To obtain an overview of the genome-scale differences among these eight groups, we performed principal component analysis and created scatter plots of the first two principal components. As expected given the established differential methylation between tissue types [25], the greatest differences were between dermis and epidermis, supporting the experimental rationale of dissociating those two tissues. There was also clear separation between samples obtained from either sun-protected or sun-exposed arm and from the face, indicating that this anatomical difference had to be considered during subsequent analysis. When a principal component analysis was repeated with just the epidermal arm samples, there was striking separation among the four groups, with the greatest difference between O-exp and Y-pro (Fig 1.2A). The sun-exposed samples obtained from the face cluster separately, however maintain the separation between old and young. Principal component analysis on dermal arm samples did not show separation related to age and exposure (Fig 1.2B), suggesting that the

number of differences would be much less than in epidermis, consistent with the previous studies[96].

Hypomethylated blocks associated with aging and sun exposure in epidermis

In previous studies comparing methylation in colorectal cancer and normal samples, large hypomethylated blocks spanning large, primarily CpG-poor regions of the genome were seen to be the major source of methylation change in cancer [28, 108]. Given the large differences found by principal component analysis within epidermal samples, we hypothesized that similar large blocks might underlie the methylation changes seen with age and sun exposure. Recent advances in analysis of 450k data through the *Minfi* package now make it possible to identify such large blocks in addition to small differentially methylated regions by applying the Bumphunter approach to probes in open sea regions using a large smoothing window [110, 113]. We began by using this block finder to identify any large blocks differentially methylated between the samples that appear most different upon principle component analysis, the epidermal O-exp and Y-pro samples. Sun exposed samples obtained from the face were included in this analysis. Identified blocks were filtered by size and the calculated family-wise error p-value. We identified 224 blocks, of which 223 were hypomethylated in the O-exp samples compared to Y-pro samples, with an average size of 443 Kb (ranging from 202 Kb to 1.3 Mb) and 9.2% reduction in DNA methylation, ranging from 5.2-16% (examples in Fig 1.3A and Fig 1.3B, list of block locations in Table 1.2). These blocks cover a total of 99 Mb of the genome. It is likely that more blocks of hypomethylation are present in these samples, however our detection is limited by the coverage of the 450k array in open sea regions.

We first sought to rule out confounding by the fact that exposed and protected samples are obtained from slightly different sites on the body (outer arm vs. inner arm). If the large scale differences observed were due to body location per se and not chronic exposure, we should detect these changes comparing the exposed and protected samples obtained from younger individuals. To address this, we performed the block finding analysis comparing Y-exp and Y-pro samples. Using the same cutoffs applied previously, we identified only 12 regions that differ, encompassing 3.8 Mb of the genome—compared to 224 regions, encompassing 99 Mb for the O-exp vs. Y-pro comparison (Table 1.3). While significant, the relatively small area involved (3.8Mb compared to 99 Mb) indicates the magnitude of block hypomethylation observed comparing O-exp and Y-pro samples cannot be attributed only to body site differences. By contrast, when comparing the O-exp and O-pro samples, we identified hypomethylated blocks of a comparable magnitude to the O-exp and Y-pro comparison, 239 blocks encompassing 100Mb, with a mean difference of 8.6% hypomethylated. While these analyses were performed without regard to individual specific changes, comparing the paired and unpaired t-statistics for each methylation location demonstrates that the unpaired analysis is conservative and only underestimates the magnitude of change present (Figure 1.4). This further indicates that only long term sun-exposure, as seen in older donors, is sufficient for hypomethylated blocks.

We also sought to determine if hypomethylation at specific regions may be linked to age alone by comparing O-pro and Y-pro samples. However, using the cutoffs applied previously, no blocks were detected. This is consistent with the lack of large scale change associated with intrinsic aging reported previously[96]. By contrast, when we compared

O-exp and Y-exp samples, we identified 33 block regions with an average of 10% difference in methylation, 32 of which are hypomethylated age, encompassing 12.7 Mb (Table 1.4). The presence of age-related hypomethylated blocks in the sun exposed body site but not the sun protected body site further demonstrates that hypomethylated blocks arise only in epidermal tissue affected by both aging and sun exposure. In contrast to the epidermis, all block finding comparisons in the data from dermal samples identified no significant blocks, consistent with the lack of differential clustering seen in the principal component analysis. Thus, hypomethylated blocks encompassing a significant portion of the genome are only observed in epidermal samples affected by both age and sun exposure. Results of all comparisons are summarized in Table 1.5.

Hierarchical clustering of all 38 epidermal samples based on methylation within identified blocks demonstrated a progressive hypomethylation with age and sun exposure: the Y-pro samples are most methylated in blocks; O-pro and Y-exp samples have lower block methylation; and O-exp samples have the lowest methylation in identified blocks (Fig 1.3C). Furthermore, this showed individual variation in the magnitude of change in blocks, with some donors showing much larger changes than others. For example, donor 8 has the lowest methylation in the sun-exposed sample across most identified blocks, while the sun exposed sample from another older individual, donor 16, is more methylated within the identified blocks. Sun-exposed samples obtained from the face (denoted in Figure 1.3C) fall with sun-exposed samples obtained from the arm, indicating that the observed hypomethylation is not specific to the dorsal forearm.

Confirmation of hypomethylated blocks by whole genome bisulfite sequencing

To confirm the presence of hypomethylated blocks related to age and sun exposure in epidermal tissue and to examine the whole genome, we obtained sun exposed and sun protected epidermal tissue from 3 additional younger donors (mean age 22 years, average Griffiths' grade 0) and 3 additional older donors (mean age 77 years, average Griffiths' grade 5) and performed whole genome bisulfite sequencing (WGBS). We generated sequencing data to a depth of 5.6x-7.2x and analyzed it using the BSmooth algorithm, which was designed for analyzing low-coverage WGBS data and has been previously demonstrated to accurately estimate methylation levels at a single-base pair resolution by borrowing information from nearby CpGs [45]. After filtering reads with low quality measures, we obtained measurements for an average of 24,873,842 CpGs per sample (average of 88.15% coverage). Bisulfite conversion, assessed using spiked in lambda phage, ranged from 99.7-99.9% (Details in Methods, Tables 1.6,1.7).

We first examined CpGs within the regions identified as blocks comparing the O-exp and Y-pro samples in 450k. All 450k O-exp/Y-pro blocks showed hypomethylation in the O-exp samples in WGBS data, with 185/224 blocks showing >5% mean hypomethylation in O-exp samples in the sequencing data, validating our finding of widespread hypomethylated blocks in O-exp epidermis. For block regions depicted in Figure 1.3A and 1.3B, methylation in WGBS samples is shown in the lower panel. Mean methylation difference from WGBS for all 450k blocks is indicated in Table 1.2. When only methylation within blocks identified in 450k data was examined, we observed separate clustering of O-exp samples from Y-pro samples, indicating these regions are sufficient to differentiate groups within sequencing data (Figure 1.5).

Density plots of measured methylation at all CpGs with sufficient coverage in all sample groups showed hypomethylation in sun-exposed epidermal samples from older donors (Fig 1.6A). Using BSmooth [45], we identified blocks within our WGBS dataset. As in 450K data, we identified large regions of change comparing the O-exp and Y-pro epidermal samples. The bisulfite sequencing analysis identified a similar mean loss in DNA methylation, 9% (ranging from 5-23%), but a much more substantial fraction of the genome than found by 450K analysis, i.e. 670 Mb (21% of the genome). The mean loss in methylation comparing the more closely related groups found similar reductions in mean methylation but fewer numbers of blocks than in 450K analysis, likely because of the smaller sample size for sequencing and thus reduced power to detect differences between the more closely related groups in the latter case (summarized in Table 1).

The observed global hypomethylation in O-exp samples is explained by CpGs within the identified blocks (Fig 1.6B and 1.6C). Thus, the results of this replication set independently confirmed our finding of hypomethylated blocks in samples affected by age and sun exposure. Examples of block regions identified in WGBS are shown in Figure 1.6D.

Hypomethylated blocks overlap with heterochromatic domains

The increased genomic coverage provided by sequencing data allowed us to compare the identified regions to other genomic domains identified through sequencing-based methods. The hypomethylated O-exp vs. Y-pro blocks in our data were seen to overlap strongly with previously reported hypomethylated blocks in human colon cancer [114] (odds ratio 8.5, $p < 2 \times 10^{-16}$). These O-exp vs. Y-pro blocks similarly showed significant overlap with Large Organized Chromatin K (lysine) methylation

heterochromatin domains (LOCKs) (odds ratio 3.4, $p < 2 \times 10^{-16}$) and with lamin-associated domains (LADs) (odds ratio 3.5, $p < 2 \times 10^{-16}$) mapped in fibroblast cell lines [29, 114]. Similar overlap of hypomethylated domains with these heterochromatin structures has been described in cancer and EBV-transformation of lymphocytes [28, 114], suggesting a functional connection between the age and exposure associated alteration in large domains of DNA methylation and processes related to cancer.

In other studies, small age-related DMRs have been found to occur preferentially at regions marked by both H3K4me3 and H3K27me3 in embryonic stem cells (bivalent domains) [85, 115]. We compared the identified blocks to regions identified as containing both H3K4me3 and H3K27me3 in human embryonic stem cells [116] and saw no enrichment (odds ratio 0.68, $p < 2 \times 10^{-16}$). Further, we note that only 17 out of 353 of the CpGs seen to correlate strongly with chronological age across multiple tissues by Horvath [89] are found within the identified block regions. This suggests that the observed large regions of hypomethylation in aged, sun-exposed epidermis represent a distinct change from the chronological age specific changes reported across other studies.

Methylation in blocks associated with clinical measures

Given the observed individual variation, we sought to determine how well block methylation levels predict clinical changes associated with skin aging. For each skin donor, a dermatologist evaluated their degree of apparent skin aging in sun-exposed and sun-protected regions using two established scales. Griffiths' photodamage age grading measures coarse and fine wrinkling, photopigmentation and yellowing [117], and showed a significant correlation with mean block methylation in sun-exposed epidermal samples ($R^2 = 0.61$, $P<0.001$) (Fig 1.7A). Much of the variation in this relationship appears to be

linked to the younger, sun-exposed samples obtained from the face, which appear to be more hypomethylated for a given age grade than the arm samples. When these samples are removed from regression, correlation is even stronger, with an R^2=0.81, p-value <0.001 (Fig 1.3D). Helfrich's photo-protected skin aging scale, in contrast, was developed for sun protected skin and measures more nuanced fine wrinkling [118]. This sun-protected scale showed far less pronounced correlation with block methylation in sun-protected samples (R^2=0.16, P=0.1) (Fig 1.7B), suggesting that the block hypomethylation correlates with clinical grading primarily in epidermal samples affected by sun exposure. Thus, methylation levels in O-exp vs. Y-pro blocks decrease with chronic exposure, varying with the degree of clinically appreciable photoaging.

To explore how genetic diversity may relate to the observed hypomethylated blocks in aged, sun-exposed skin, we examined signal at 65 single nucleotide polymorphism (SNP) probes included on the 450k array. None of the 65 probes showed a significant difference (adjusted p-value<0.05) between samples from older and younger individuals, even though these probes were chosen to discriminate genetic structure[109]. Additionally, we compared identified block regions to the genetically controlled methylation clusters (GeMes) mapped in a recent work using data from two large Caucasian cohorts[119], and genetically controlled CpGs identified using a large cohort of female twins through the MuTHER study [120]. 84 of the identified O-exp vs. Y-pro blocks contain GeMes, however the GeMes account for only 6.6 Mb of the identified 99 Mb, indicating that a large area is involved in which methylation is not directly linked to genotype. Similarly, of the 13,378 probes involved in our identified O-exp vs. Y-pro blocks, only 3,236 were linked to genotype. GeMes within the identified blocks are

denoted in Table 1.2. Furthermore, if the effect observed were genetically driven, hypomethylated blocks would be present in both samples obtained from the involved individuals, but they are found only in older, sun-exposed samples.

Given previously reported relationships between DNA methylation and body mass index (BMI) and smoking [120-122], we compared methylation levels within the identified blocks to BMI and between smokers and non-smokers. We observed no significant relationship between BMI and block methylation ($R^2 = 0.02$, p=0.56) or smoking status for donors for which this information was available ($R^2 = 0.20$, p=0.32) (Fig 1.8A and 1.8B).

Differentially expressed genes within hypomethylated blocks

To determine the relationship between hypomethylated blocks observed in epidermal samples with sun exposure and aging, we obtained RNA from epidermal samples from a subset of the donors, including 3 Y-pro epidermal samples and 4 O-exp epidermal samples (individual donors used denoted in Table 1.1). Expression data was obtained using Affymetrix arrays. When we compared O-exp and Y-pro samples using Gene Set Enrichment Analysis (GSEA) [123], 88 pathways were identified as significantly enriched based with an FWER < 0.05. Intriguingly, this analysis identified 3 pathways associated with UV exposure *in vitro*, as well as multiple pathways linked to cell cycle and proliferation (full results in Table 1.8). To understand the relationship between sample groups, we used *limma* to compare each group of samples. When comparing the O-exp and Y-pro samples, we observed more probes with upregulation in the O-exp samples (Fig 1.9A and 1.9B). When we considered the intrinsic aging comparison (O-pro vs Y-pro), we did not observe this trend (Fig 1.9C and 1.9D). When

we considered the Y-exp vs Y-pro comparison, we noted a similar trend towards upregulation in the sun exposed samples (Fig 1.9E and 1.9F). This indicates that, unlike the pattern observed with methylation, the global changes in gene expression are driven by the exposure/body site differences, not the combination of age and sun exposure.

Given our observation of widespread methylation changes within the epidermal samples, we sought to use our gene expression data to confirm that no large shifts in cell type composition may be confounding our analysis. The second most common cell type in the epidermis, melanocytes (~3% of cells), can be distinguished from keratinocytes based on the presence of MITF and SOX10 [124]. To determine if large changes in melanocyte levels are present in our samples, we examined expression probes measuring expression of these markers (3 measuring MITF and 2 measuring SOX10). None of the examined probes showed a significant difference in expression comparing O-exp and Y-pro samples, and none showed a log fold change greater than 0.5 (Fig 1.10).

In order to determine how patterns of expression relate to the identified hypomethylated blocks, data was normalized using fRMA and probes were classified as expressed or unexpressed using gene expression barcode [125]. When expression data was compared to the locations of photoaging blocks, probes within blocks were less likely to be expressed than those outside (OR 0.38, $p<0.001$), consistent with the observed overlap between blocks and heterochromatic regions.

Blocks hypomethylated with sun-exposure and aging are hypomethylated in squamous cell carcinoma

Having observed an overlap between our chronic exposure blocks and hypomethylated domains observed in colon cancer, we sought to determine the relevance

of our identified domains to a cancer that develops in epidermis with chronic sun exposure, squamous cell carcinoma (SCC). We obtained 7 SCC tissue samples and 6 normal skin samples from the same body sites as the SCC samples and analyzed DNA methylation using the 450k array. Strikingly, we observed hypomethylation of SCC samples compared to normal samples when examining probes within the identified chronic exposure blocks (Fig 1.11A), which is not seen when examining probes outside of these regions (Fig 1.11B). This difference is seen even when examination is limited to probes within the constitutively methylated open sea probes (Fig 1.11C and 1.11D). For 221 out of 223 identified hypomethylated blocks, SCC samples had lower mean methylation than normal samples (difference in mean methylation noted in Table 1.2). Clustering of this data based on mean methylation within identified photoaging blocks distinguishes most SCC from normal samples (Fig 1.11E).

Methylation age correlates strongly with chronological age

In a recent work, Horvath built an age-prediction model using methylation at 353 cytosines probed on the 450k array which allows calculation of a "methylation age" for any given sample[89]. Horvath reports these methylation ages are highly correlated with chronological age using a wide range of tissue samples. We applied Horvath's model to calculate methylation age for our samples to determine how the widespread hypomethylation we observe in O-exp epidermis relates to methylation age. We observed a strong correlation between chronological age and methylation age for both dermis (R^2=0.91, p<2.2e-16) and epidermis (R^2=0.83, p<1.1 e-15) (Fig 1.12).

Small differentially methylated regions overlap with polycomb targets

We next used our 450k data to identify small differentially methylated regions. We used the method of "bump hunting," which identifies genomic regions in which methylation levels at consecutive measured locations are associated with the outcome of interest. That method was initially developed for the CHARM array based methylation method but was adapted for the Minfi package for the 450K array [110, 113]. Significance testing was performed by permutation analysis. Comparing O-exp and Y-pro epidermal samples, we identified 166 small DMRs with at least 10% difference in methylation and an adjusted family wise error rate <0.1. These included 90 hypermethylated and 76 hypomethylated DMRs (list in Table 1.9). The average size of these DMRs was 460 bp, and the average change in methylation was 37% (ranging from 23-64%), indicative of local regional effect on DNA methylation, similar to what is seen in tissue and cancer DMRs.

As we considered with the block regions, we asked what changes were specific to age or sun exposure. Similarly to the block analysis, we identified a considerably smaller number of DMRs specific to sun exposure (30 significant DMRs found comparing Y-exp and Y-pro samples vs. 166 comparing O-exp and Y-pro) and specific to age (2 significant DMRs found comparing O-pro and Y-pro samples vs. 166 comparing O-exp and Y-pro). Within dermal samples, we again identified fewer regions of change than in the epidermis: 9 significant O-exp and Y-pro DMRs were found, with even fewer sun-specific and age-specific DMRs. Complete DMR counts from all comparisons are shown in Table 1.10. These results suggest that small DMRs, like blocks, require both age and

sun exposure and occur predominantly in the epidermis, although the cumulative size of these DMRs was less than 1% of the area covered by large blocks on 450k.

In other studies, small age-related DMRs have been found to occur preferentially at regions marked by repressive chromatin, bivalent domains, and targets of Polycomb Repressive Complex 2 [85, 115, 126, 127]. To determine if this overlap is found in the identified small DMRs, we compared the identified regions to regions marked by H3K4me3 and H3K27me3 in normal human epidermal keratinocytes (NHEK cells) as identified by the ENCODE project [128]. To control for the background enrichment from the 450k array, we determined the significance of identified overlaps by generating length and probe number matched random regions from all array probes. We called an overlap significant if there was more overlap between identified DMRs and chromatin mark than >95% of randomly generated region sets. We observed a significant overlap between hypermethylated O-exp and Y-pro DMRs and regions marked by H3K27me3 in NHEK cells, but not regions marked by H3K4me3. Hypomethylated OE-YP DMRs did not overlap significantly with either mark.

Discussion

In summary, a comprehensive analysis of DNA methylation in aging and sun-exposed skin reveals widespread DNA hypomethylation of large blocks that overlap heterochromatin domains and nuclear lamin-associated domains. The greatest difference was between older, sun exposed and younger, sun protected epidermal samples, with indicating that both age and sun exposure status is necessary for widespread hypomethylation. We further observe these same regions to account for the majority of the hypomethylation in squamous cell carcinoma samples, suggesting a possible link

between widespread hypomethylation in cancer and the changes observed in normal epidermis.

The most unexpected result of our analysis was the genome-scale differential methylation seen in epidermal samples affected by both age and exposure, as evidenced by the distinct clustering in principal component analysis. While most population studies of aging and environmental exposures have focused on methylation changes in single CpGs or small differentially methylated regions [85, 89, 102], here we identify changes across large block regions encompassing ~20% of the genome. This extent of altered methylation suggests that the change associated with chronic environmental stress should be studied on a genome scale, rather than only focusing on changes in DNA methylation in discrete regions. While our sample size relatively small, the large scale change identified is informative for future study. The overlap of our identified blocks with other structural domains, including LADs and LOCKs, further supports the idea that the observed changes in DNA methylation are indicative of alteration to the epigenomic structure. Future studies of chronic exposure should examine methylation at a similar, genome, scale to determine if the changes identified in our sample set are more broadly relevant.

We chose to examine skin in this study because it offers a relatively homogenous system with regard to cell type for studying aging and exposure. Once separated, the epidermis in particular consists of 90-95% keratinocytes in both sun-exposed and sun-protected body sites[106]. The other main cell types in the epidermis, melanocytes and Langerhans cells, account for a small percentage of cells—there are approximately 36 keratinocytes per melanocyte and approximately 53 keratinocytes per Langerhans

cell[66]. While small increase in melanocyte density in sun-exposed body sites and small decreases in melanocyte density with age have been reported, these are not of the correct pattern or sufficient magnitude to confound the observed blocks of hypomethylation: comparing sun-exposed and sun-protected body sites from young and old individuals shows up to a twofold increase in melanocytes in heavily sun-exposed sites and a decrease in melanocyte density by 6-8% per decade in both sun-exposed and sun-protected locations[129]. The consistent level of expression of melanocyte markers within our sample set confirms that no unexpected large changes in melanocyte composition are present in our study. As melanocytes account for ~3% of cells present, a doubling or complete loss of these cells could account for up to a 3% change in methylation, smaller than the magnitude of change observed in blocks. Further, as we demonstrate, the observed hypomethylated blocks are not present comparing sun exposed and sun-protected body sites in young individuals, so they can not be attributed to the increased percentage of melanocytes present in sun exposed body sites of both young and old individuals. The age related decrease in melanocytes would amount to approximately a 1% change in cell composition over 4 decades (32% decrease in 3% of cells), which is again, not sufficient to account for the changes we observe. Consistent with the very small changes in cell composition reported in the literature, analysis of gene expression data from our samples indicates no significant changes in melanocyte marker genes.

In the recent literature, multiple methods have been developed to in silico correct for cell type heterogeneity, including SVA, EWASher and refFreeEwas[130-132]. All of these methods assume that the signal between cases and controls is small compared to the effect of cell type heterogeneity and that any large scale change identified in principle

component analysis is due to cell composition changes. Using these methods will therefore remove any large scale signal between cases and controls, even if these differences are not due to cell composition effects. An example is given by the EWASher paper in which applying the methodology to comparison of colon cancer tumors and normal colon leaves only 2 CpGs identified as significantly differentially methylated in cancer, contradicting 25 years of literature on this disease[131], and acknowledged by the authors themselves to be a significant blind spot of such correction methods [110, 133]. Despite the significant body of evidence supporting the absence of large scale cell composition changes in our samples, the methylation changes we describe encompass a large portion of the genome and are large enough to affect clustering (Figure 1), so it is not surprising that the application of these algorithms to our data removes the widespread hypomethylation observed in the old, sun exposed samples. However, given the substantial evidence indicating no large changes in cell composition with aging and sun-exposure, we attribute this to the degree of differential methylation within the keratinocyte population and not changes in cell type composition.

The identified blocks are distinct from the single CpGs used to calculate "methylation age" by Horvath and the small hypermethylated DMRs overlapping bivalent chromatin domains identified as markers of chronological age in other studies [85, 89, 115]. We see no evidence of hypomethylation within our blocks or within 450k open sea probes associated chronological aging in publically available data from adipose tissue or peripheral blood [102, 120], reinforcing our finding that block hypomethylation in epidermal tissue occurs only with chronic exposure, not aging alone. Heyn et al. report widespread hypomethylation with age in CD4+ T cells from WGBS of cells sorted from

one newborn and one centenarian [94], indicating that large blocks of hypomethylation may be relevant in other models of aging, however we see only a moderate overlap of our identified blocks with hypomethylated CpGs in their study (OR= 1.77) and the use of single samples in that study makes it difficult to draw further conclusions about the regions involved.

A functional role of altered block methylation is supported by the correlations between block methylation and clinical measures of skin aging. Our analysis of the relationship between hypomethylated blocks and gene expression is limited by the low number of samples available for gene expression analysis. In studies of cancer and EBV transformation, large publically available expression datasets have allowed identification of a clear relationship between hypomethylation and variability of expression of genes within blocks[28, 114]. Another recent study linked loss of methylation in gene poor regions with activation of distal enhancers, indicating that loss of methylation may have functional consequences beyond regulation of nearest genes[134]. Consistent with an environmental exposure-driven effect, the magnitude of age and sun related change was greater in epidermal samples compared to the dermal samples. These results suggest that molecular aging and UV exposure studies focused on dermis or full thickness biopsy may miss the relevant epigenetic effects [135].

The occurrence of blocks of hypomethylation related to age and sun exposure in regions previously seen to be hypomethylated in cancer is highly intriguing. Similar blocks identified in colon cancer samples were seen to develop very early in cancer progression [28], suggesting epigenetic reorganization may be an early event in carcinogenesis. While the blocks observed in our study encompass a smaller portion of

the genome and have a smaller magnitude of change, their occurrence in non-malignant tissue suggests that epigenomic change may be initiated by exposures, rather than only during oncogenesis. Our observation in this work, that blocks identified as hypomethylated with chronic exposure in non-malignant tissue are hypomethylated in squamous cell carcinoma, while most other sites are not, suggest that these changes may occur prior to malignancy. Our identification of this subset of the "cancer-blocks" that become hypomethylated with chronic exposure offers a potential target for chemoprevention strategies.

We note that these results suggest that an early change quantitatively related chronic exposure in skin is a large-scale epigenomic alteration that was missed by earlier studies focused on individual genes or loci such as CpG islands. Intriguingly, a recent study by Cruickshanks et al. identifies similar large regions of hypomethylation in cells cultured to replicative senescence [95], linking such loss of methylation to cell turnover and cell stress in culture. Further, Takeuchi et al. recently demonstrated that repeated exposure of aged cells to UVA induces expression of progerin, the mutated form of Lamin A seen in Hutchinson-Guilford progeria, and abnormal nuclear morphology [136]. This is intriguing given the observed overlap between our blocks and lamina-associated domains. While speculative, a general model consistent with these data is that repeated stress and cell turnover leads to progressive destabilization of heterochromatin along the nuclear membrane, with attendant changes in DNA methylation. These changes could serve both as a molecular epigenetic clock of environmental exposure and a potential target of increased risk. They fit well an age-associated epigenetic drift hypothesis [104]. These results may have therapeutic implications. For example, future experiments should

be performed to determine whether modifiers of sun damage such as retinoic acid, or specific inhibitors of laminopathies such as lonafarnib [137] might modify the development of the epigenomic alterations described here. Finally, we emphasize that the changes observed here may be restricted to the specific sample types and method of analysis used here.

Experimental Procedures

Human subjects and tissue samples: Twenty-six healthy volunteers took part in the study. Samples from twenty of these donors were used for 450k analysis, and samples from the additional six donors were used for WGBS. All participants provided written informed consent, and this study was approved by the Johns Hopkins Medicine Institutional Review Board (IRB# NA_00041408 for sun-exposed and sun-protected samples, NA_0036868 for SCC samples). All human experimentation was performed in accordance with the Declaration of Helsinki. Exclusion criteria included use of topical medications within 2 weeks of sampling, active or dormant skin conditions, and pregnancy. To control for pigmentation, only Caucasian subjects were selected. For each subject, a dermatologist evaluated the degree of apparent skin aging using two established scales: Griffiths' photodamage age grading measures coarse and fine wrinkling, photopigmentation and yellowing [117] while Helfrich's photo-protected skin aging scale, in contrast, was developed for sun protected skin and measures more nuanced fine wrinkling [118]. Paired punch biopsy samples, 4 mm in diameter, were collected under local anesthesia from the outer forearm or lateral epicanthus (sun-exposed area) and upper inner arm (sun-protected area). Immediately after removal, samples were washed in PBS and transferred to DMEM (Gibco) containing dispase (2

U/mL). Following dispase treatment overnight at 4°C, epidermis and dermis were separated, flash frozen and stored at -140°C. SCC containing tissue was obtained from ten previously diagnosed patients (8 females and 2 males; median age 67yrs) undergoing either Mohs micrographic or excisional surgery for treatment of their skin cancer. During repair of the surgical defect, standing cone deformity "dog ear" tissue was retained and submitted as normal control tissue. Samples were OCT embedded, frozen and stored at -80°C.

Histology: Dispase separated, flash frozen epidermal sections were cut into 5 um sections and stained with hematoxylin and eosin. Sections were imaged using a Vectra Imaging system.

DNA isolation: DNA was isolated from biopsy samples (epidermis and dermis) and SCC samples using the MasterPure DNA Purification Kit (Epicentre) according the manufacturer's protocol.

450k Array: DNA was quantified using Quant-iT Picogreen Reagent (Invitrogen) according to manufacturer's instructions. 1 μg of DNA was bisulfite treated using the EZ DNA methylation kit (Zymo Research) according to the manufacturer's specifications for the 450k array. Converted genomic DNA was eluted in 22 μl of elution buffer. DNA methylation level was measured using Illumina Infinium HD Methylation Assay (Illumina) according to the manufacturer's instructions.

450k Analysis: All analyses were performed using R 3.0.3. Raw intensity files were obtained and processed using the Minfi package to obtain methylation ratios (Beta values). Samples were normalized using the Illumina preprocessing method implemented

in Minfi. We applied multiple quality control measures to remove questionable arrays or probes. We examined 450k array control probes to assess many measures of assay efficiency and calculated median methylated and unmethylated measurements for each sample. We removed probes that had an annotated SNP (dbSNP137) at the single base extension or CpG site (17541 probes removed).

To identify blocks, we used the block finder as described elsewhere [110]. Briefly, we applied the Bumphunter approach to the open sea probes using a large smoothing window. The estimates were thresholded based on a 5% difference in methylation beta values. Blocks were filtered to include only those >200 kb. Significance was assigned based on permutation testing; a cutoff of adjusted p-value (Family wise error rate) < 0.05 was used.

To identify small DMRs, we used the bump hunting technique as previously described [138]. Estimated differences were controlled for sex and body site (face or arm) in the linear model. The estimates were thresholded using a 0.1 difference in beta values (~10% difference in methylation). Significance was assigned based on permutation testing; a cutoff of adjusted p-value (Family wise error rate) < 0.1 was used.

To examine overlap with histone marks, we downloaded ChIP-Seq peaks from NHEK cells (Lonza CC-2501) generated by the Bernstein-Broad group for the ENCODE project (GEO accession GSM733701, GSM733720) and determined how many of our identified DMRs overlap with each set of marks. To assess the significance of overlap with the identified DMRs, we generated random regions from 450k probes with the same width and probe number as the identified regions, determined how many of these random

regions overlap with each histone mark and repeated this procedure 1000 times. The DMR list was considered to significantly overlap with a histone mark if it contained more overlaps than >95% of randomly generated lists.

Affymetrix microarray expression analysis: RNA was isolated from epidermal samples using Trizol Reagent (Life Technologies) according to manufacturer's instructions followed by clean up using the RNeasy kit (Qiagen) following the RNA cleanup protocol. Genome-wide gene expression analysis was done using Affymetrix U133 Plus 2.0 microarrays according to Affymetrix's specifications. Data were normalized using fRMA as previously described [139] expression was determined using gene expression barcode [125]. Probes were classified as expressed if mean expression Z-score in at least one group was >2.54. Differential expression was determined using limma [140]. Gene Set Enrichment Analysis was performed using the gene pattern suite GSEA module[141]. Significance was assessed using 1000 permutations of gene sets.

WGBS libraries: Bisulfite sequencing libraries were constructed using the Illumina TruSeq DNA Library Preparation kit protocol with the following modifications. 10 ng of unmethylated lambda DNA were added to one microgram of genomic DNA prior to shearing in order to monitor bisulfite conversion efficiency. After shearing, end repair was performed using a modified protocol to prevent introduction of non-genomic cytosines by using only dATP, dGTP and dTTP nucleotides with a mixtures of Klenow DNA polymerase, T4 DNA polymerase and T4 Polynucleotide kinase. After purification, samples were bisulfite converted and purified using Zymo EZ DNA Methylation Gold. Bisulfite converted libraries were amplified using a mixture of Uracil insensitive

polymerases, Denville Choice Taq and Agilent Pfu. Samples were amplified for 10 cycles of PCR.

WGBS Analysis: All analyses were performed using R 3.0.1. To process sequencing data, we ran the BSmooth [45] bisulfite alignment pipeline (version 0.4.5-beta) on the 100-by-100 bp HiSeq 2000 paired end sequencing reads obtained for each sample, using Bowtie2 version 2.0.1 [142] and the hg19 build on the human genome as well as the genome for lambda phage. Table 1.6 summarizes the alignment results. After alignment, BSmooth was used to extract read-level measurements, summarized in Table 1.7. We filtered out measurements with mapping quality <20 or nucleotide base quality <10 and we removed measurements from the 5' most 10 nucleotides of both mates. BSmooth was used to sort read-level measurements by genomic coordinates and compile a summary table.

Next, BSmooth was used to identify large hypomethylated blocks as described in detail previously [28, 45, 114]. CpGs with coverage of 2 or greater in each sample group (O-exp, O-pro, Y-exp, Y-pro) were included in the analysis. We used the same cutoffs used in studies of cancer, specifically a t-statistic cutoff of -2, 2. We estimated variance based on the younger, sun-protected samples. Identified blocks were filtered to include only blocks >10kb and with a mean difference of >5%.

Data Availability: Array and sequencing data are available in GEO under accession number GSE52980.

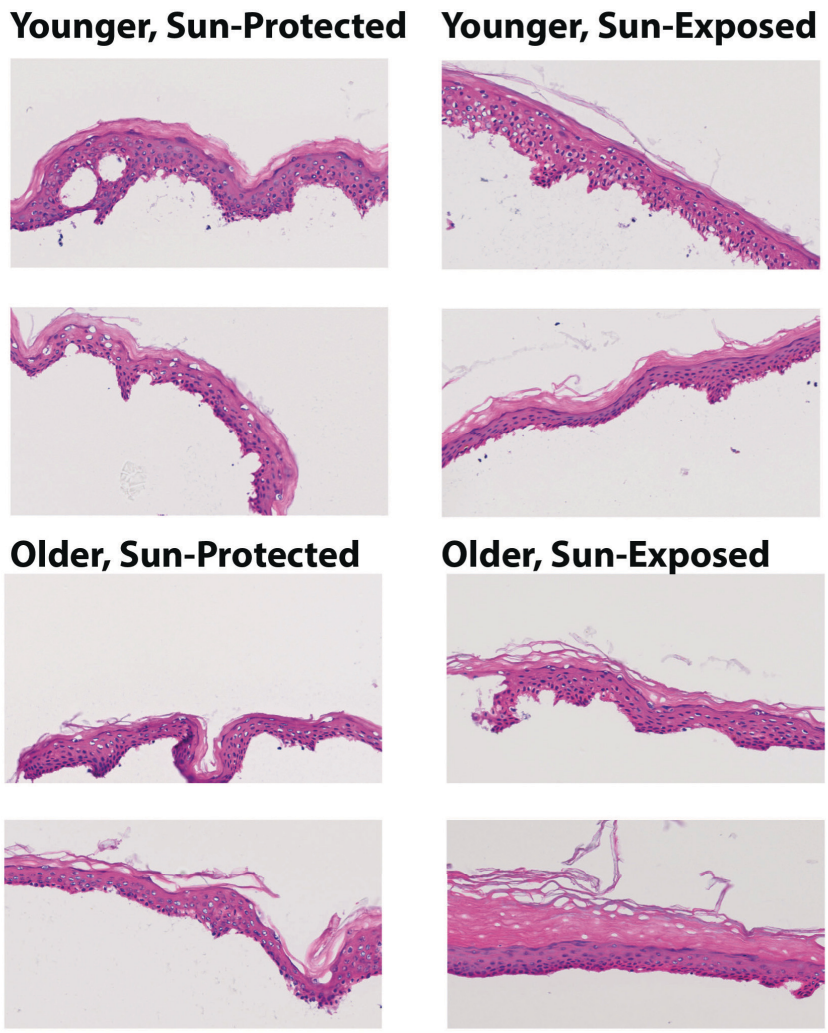


Figure 1.1: Hemotoxylin and eosin staining of dispase separated epidermis.

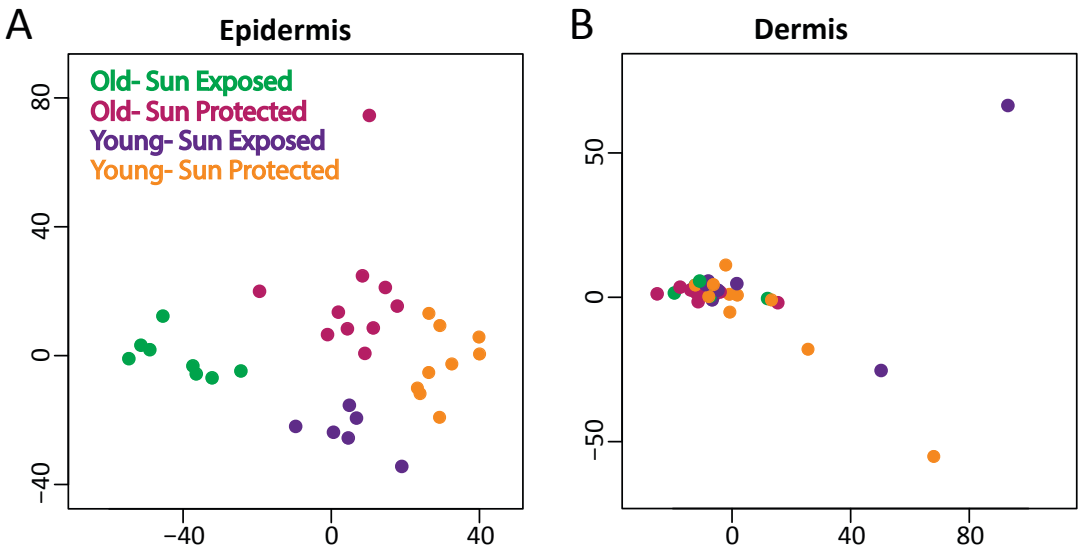


Figure 1.2: DNA methylation in epidermal but not dermal samples clusters by age and sun exposure.

(A) In epidermis, DNA methylation segregates old vs. young individuals, and also segregates sun-exposed and sun-protected anatomical regions, shown by multidimensional scaling of pairwise distances derived from methylation levels assayed on the HumanMethylation450 Beadchip. (B) In dermis, DNA methylation does not segregate samples by age or anatomical region, shown by multidimensional scaling of pairwise distances derived from methylation levels assayed on the HumanMethylation450 Beadchip.

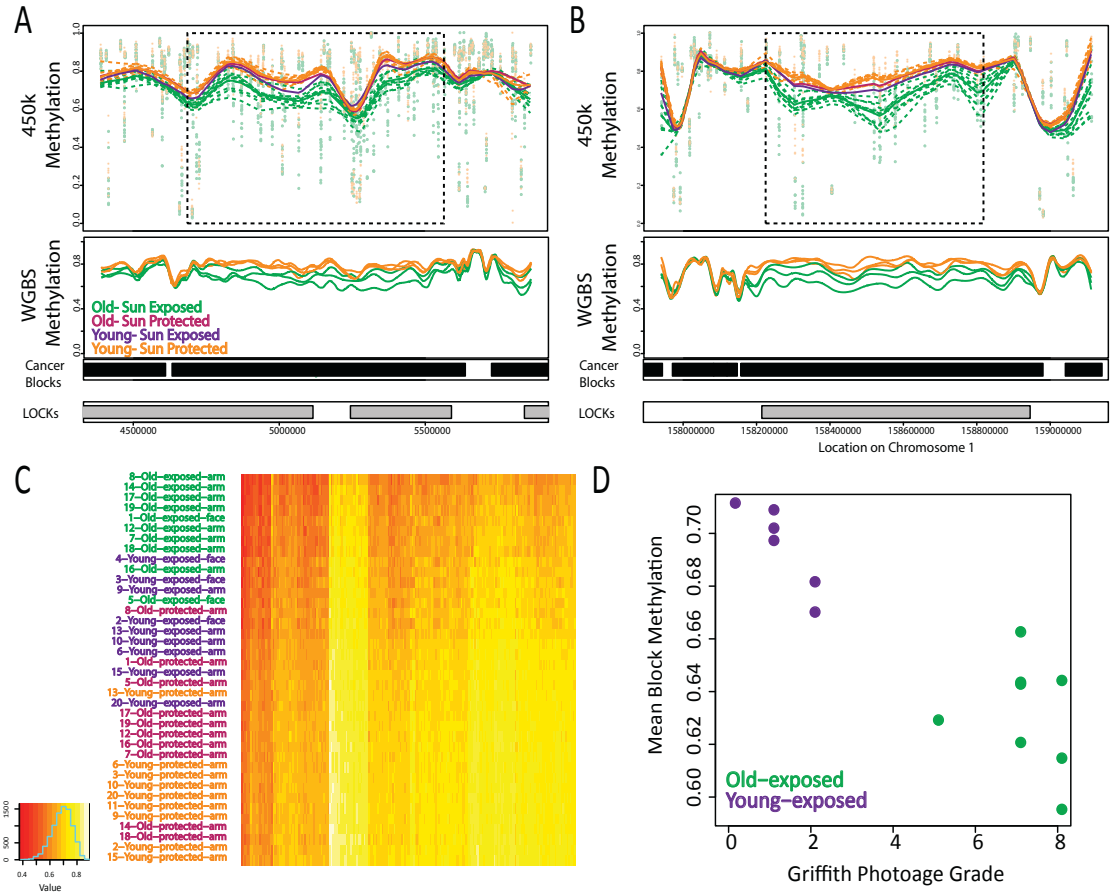


Figure 1.3: Block hypomethylation progresses from younger sun-protected to older sun-protected to younger sun-exposed to older sun-exposed tissue.

(A) Example of a region identified as a block comparing older sun-exposed to younger sun-protected epidermal samples using 450k data. Top panel: Shown are methylation beta values (methylated signal/total signal) for "collapsed" measurements of methylation from open sea probes in 450k data. These are methylation averages for each 1500 bp open sea region calculated as part of the *Minfi*'s "block finder" algorithm. The points represent individual samples at each location, dotted lines show smoothed measurements across the region for each individual and solid lines represent the smoothed average for each group. The box demarcates the block identified using *Minfi*. Bottom panel: shown are smoothed methylation beta values from WGBS data within the regions identified in 450k analysis. Horizontal bars indicate the locations of hypomethylated blocks identified previously in cancer and heterochromatin LOCKs [28]. (B) Example of a region identified as a block comparing older, sun-exposed and younger, sun-protected epidermal samples using 450k data, plotted as in A. (C) Heatmap showing mean block methylation in all blocks identified comparing O-exp and Y-pro epidermis. Samples are ordered by mean methylation, and blocks are ordered by mean difference in methylation between O-exp and Y-pro samples. Red/yellow indicate lower/higher mean methylation levels, respectively. (D) Relationship between block methylation and Griffiths' photoage grade. Mean methylation within all blocks identified comparing O-exp and Y-pro epidermis, versus Griffiths' photoage grade assigned to the sample donor.

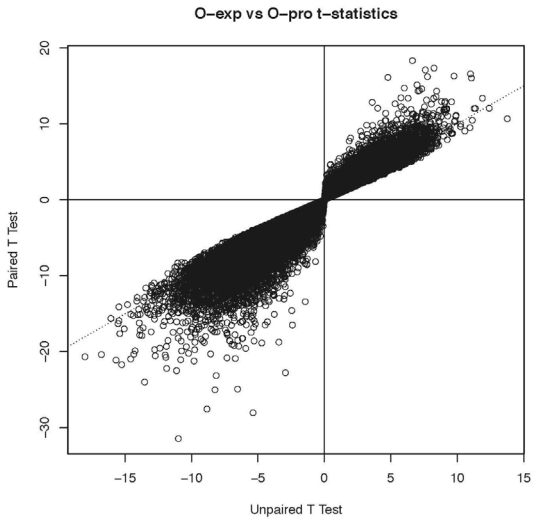


Figure 1.4: The unpaired analysis is conservative relative to a paired test.

Shown are the t-statistics calculated fro each probe comparing sun-exposed and sun-protected samples from older individuals using a paired t-test versus those calculated using an unpaired t-test.

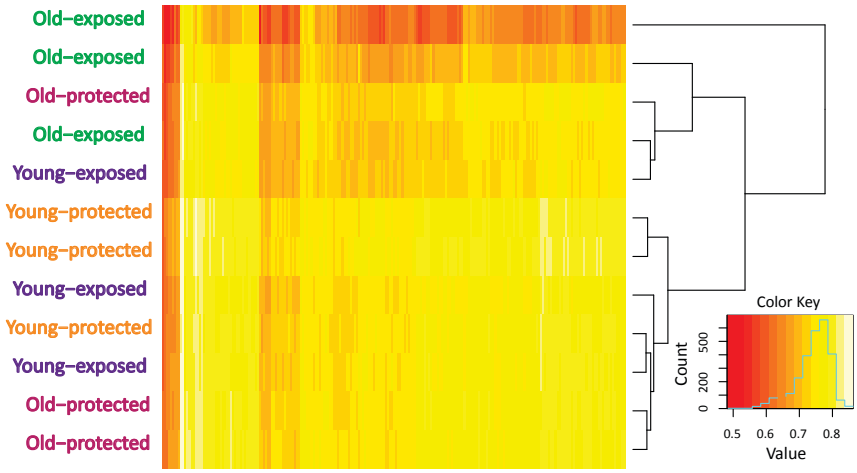


Figure 1.5: Hypomethylated blocks identified in 450k data separate O-exp and Y-pro samples analyzed using WGBS.

Heatmap showing mean methylation from WGBS data in regions identified as blocks comparing O-exp and Y-pro epidermis in 450k data. Samples and blocks are ordered by hierarchal clustering. Yellow/red indicate higher/lower methylation, respectively.

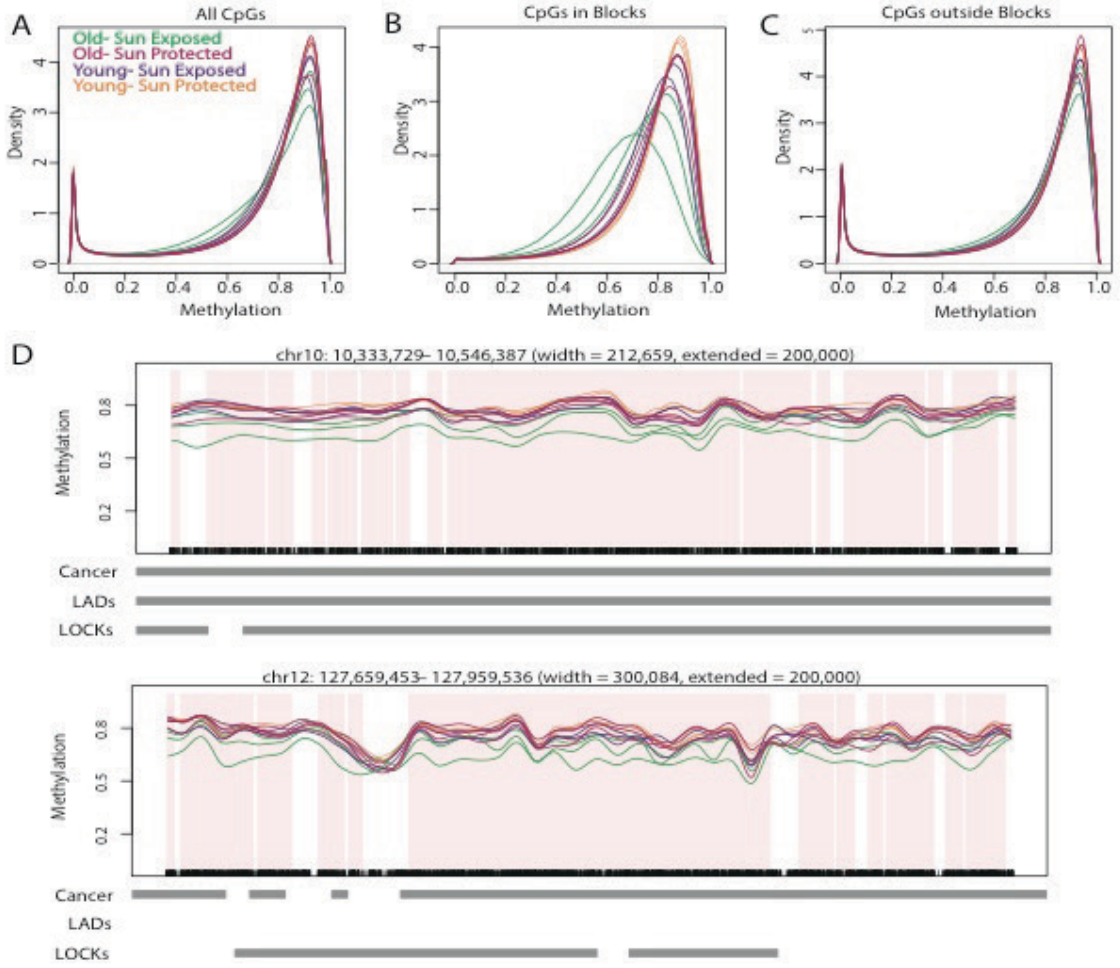


Figure 1.6: Hypomethylated blocks in O-exp samples replicated by whole genome bisulfite sequencing.

(A-C) Hypomethylation is enriched in blocks. Distribution of high-frequency smoothed methylation values from CpGs with sufficient coverage from WGBS for (A) all CpGs, (B) CpGs in blocks, and (C) CpGs outside of blocks. (D) Examples of blocks identified by WGBS. Shown are smoothed methylation values within blocks identified comparing O-exp and Y-pro epidermis (pink).

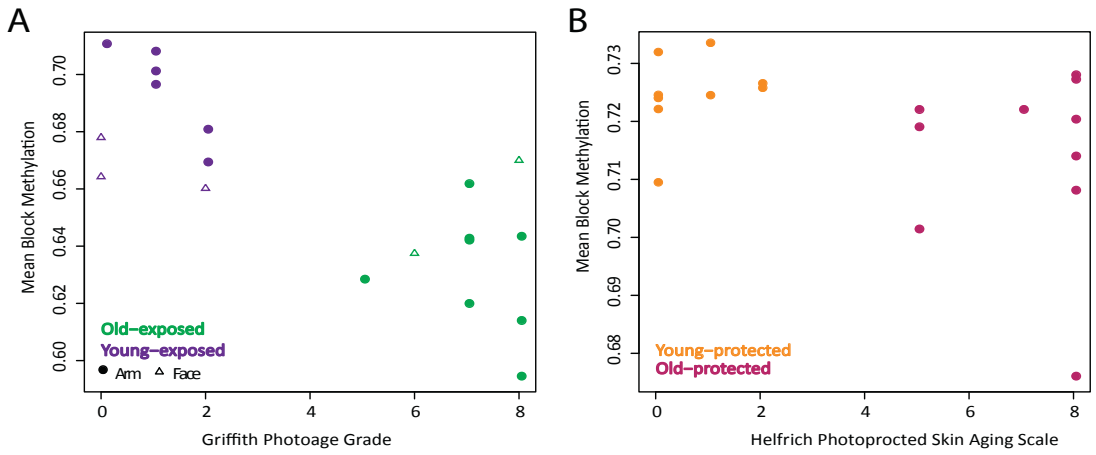


Figure 1.7: Methylation in sun exposed samples correlates with sun exposed age grade.

Shown is mean methylation within blocks identified comparing O-exp and Y-pro epidermis for each sun-exposed epidermal sample versus Griffths' photoage grade assigned to sample donor. (B) Methylation in photoprotected samples does not correlate with photoprotected age grade. Shown is mean methylation within blocks identified comparing O-exp and Y-pro epidermis for each sun-protected epidermal sample versus Helfrich photoprotected age grade assigned to sample donor.

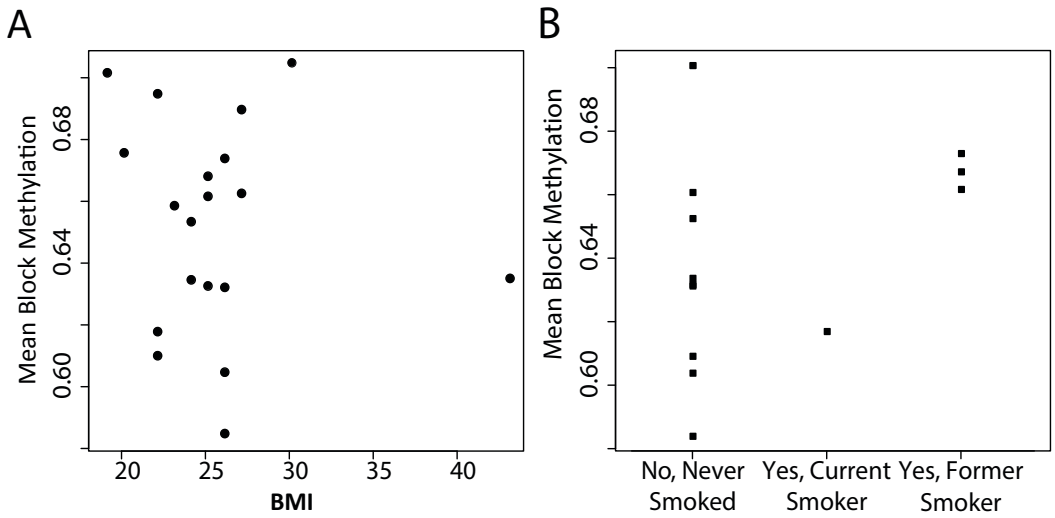


Figure 1.8: Mean block methylation does not correlate with BMI or smoking status.

(A) Mean methylation within blocks identified comparing O-exp and Y-pro epidermis for each sun-exposed epidermal sample versus donor BMI. (B) Mean methylation within blocks identified comparing O-exp and Y-pro epidermis for each sun-exposed epidermal sample versus smoking status.

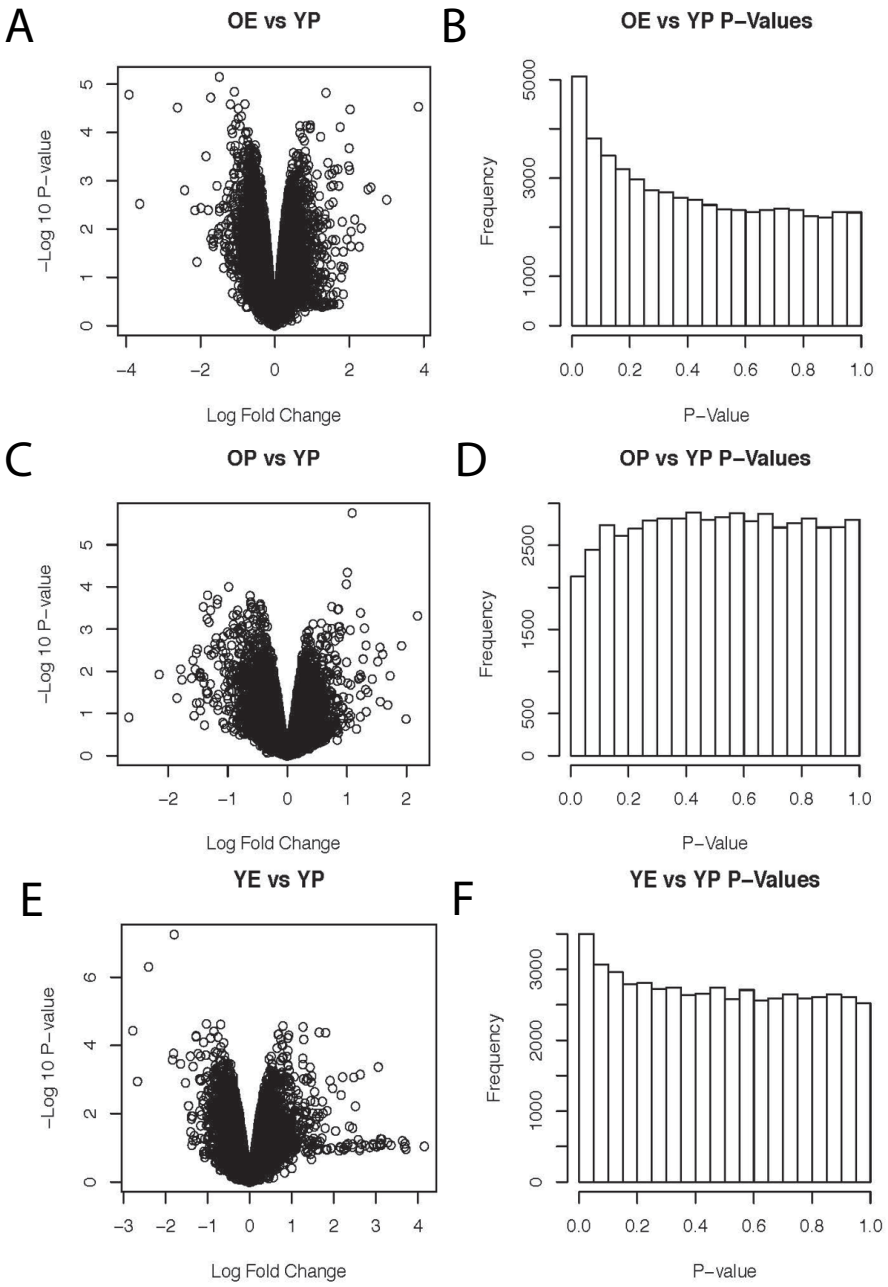


Figure 1.9: Differential gene expression in epidermal samples with sun exposure and aging.

A) Volcano plot comparing significance and magnitude of differential expression for the O-exp vs Y-pro comparison. B) Distribution of p-values for the O-exp vs Y-pro comparison. C) Volcano plot comparing significance and magnitude of differential expression for the O-pro vs Y-pro (Intrinsic Age) comparison. D) Distribution of p-values for the O-pro vs Y-pro (Intrinsic Age) comparison. E) Volcano plot comparing significance and magnitude of differential expression for the Y-exp vs Y-pro (Exposure) comparison. F) Distribution of p-values for the Y-exp vs Y-pro (Exposure) comparison.

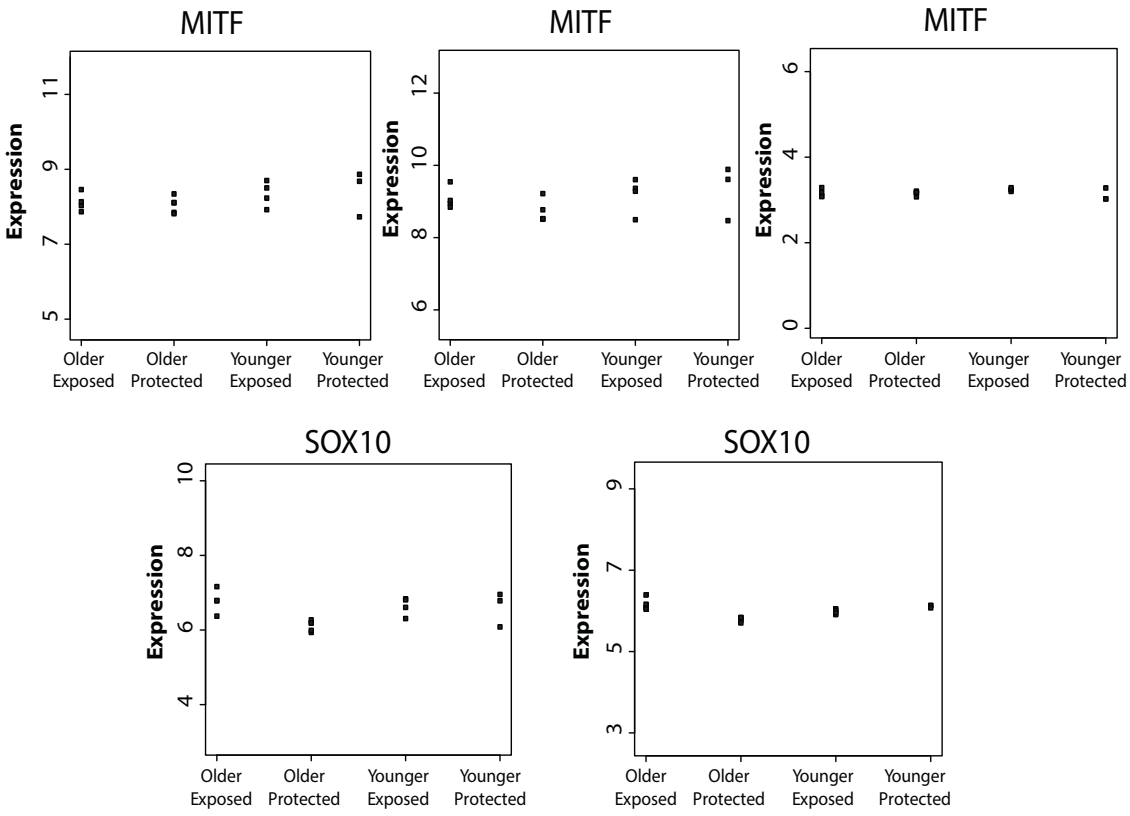


Figure 1.10: Expression of melanocyte markers is not significantly different with aging and sun exposure status.

A) Log2 expression of melanocyte marker, MITF, measured by 3 Affymetrix probes versus age and sun-exposure status for epidermal samples. B) Log2 expression of melanocyte marker, SOX10, measured by 2 Affymetrix probes versus age and sun-exposure status for epidermal samples.

Table 1.1: Donor Characteristics

Ethnicity	Biopsy Sites	Griffiths	Helfrichs	Age	BMI	Method
Caucasian	left lateral eye, UIA (Upper Inner Arm)	6	5	61	43	450K
Caucasian	left lateral eye, UIA	0	0	22	20	450K
Caucasian	left lateral eye, UIA	0	0	28	27	450K
Caucasian	right lateral eye, UIA	2	0	34	23	450K
Caucasian	left lateral eye, UIA	8	8	69	25	450K
Caucasian	right DF (Dorsal Forearm), UIA	1	0	29	22	450K
Caucasian	left DF, UIA	7	5	70	25	450K
Caucasian	left DF, UIA	8	8	72	26	450K
Caucasian	right DF, UIA	1	1	21	27	450K
Caucasian	right DF, UIA	2	2	28	25	450k
Caucasian	left DF, UIA	3	2	30	23	450k, Expression
Caucasian	left DF, UIA	7	8	74	26	450k
Caucasian	left DF, UIA	2	0	30	26	450k, Expression
Caucasian	right DF, UIA	8	8	83	26	450k, Expression
Caucasian	left DF, UIA	1	1	25	19	450k, Expression
Caucasian	left DF, UIA	7	7	65	24	450k, Expression
Caucasian	left DF, UIA	7	8	>90	22	450k, Expression
Caucasian	right DF, UIA	8	8	84	24	450k, Expression
Caucasian	left DF, UIA	5	5	65	22	450k, Expression
Caucasian	left DF, UIA	0	0	20	30	450k, Expression
Caucasian	left DF, UIA	0	0	25	22	WGBS
Caucasian	left DF, UIA	0	0	23	25	WGBS
Caucasian	left DF, UIA	0	0	18	26	WGBS
Caucasian	left DF, UIA	6	6	75	34	WGBS
Caucasian	left DF, UIA	4	6	74	25	WGBS
Caucasian	left DF, UIA	5	8	83	26	WGBS

Table 1.2: OE-YP Blocks

Chr	Start	End	FWER	OE vs. YP	SCC vs. Normal Samples	WGBS Diff
chr14	101071136	101970684	0	8.25%	7.01%	6.76%
chr1	2565265	3315330	0	8.66%	6.72%	8.64%
chr11	4686966	5566461	0	8.04%	9.11%	9.20%
chr11	55110223.5	56185228	0	9.03%	8.50%	11.73%
chr14	105974213	106444921.5	0	9.28%	7.38%	7.54%
chr1	158223723.5	158817808	0	11.22%	10.49%	11.12%
chr1	248042285	248685382	0	8.46%	5.44%	8.76%
chr11	131672204	132031541	0	8.88%	3.80%	10.64%
chr7	136563350	137139113	0	8.37%	7.85%	10.98%
chr12	97433383	98577192	0	9.50%	6.89%	10.03%
chr13	53595003	54943816	0	8.67%	5.66%	10.21%
chr7	77703597	78982418	0	9.05%	6.44%	9.14%
chr4	16504511	16883525	0	11.45%	8.06%	9.81%
chr5	156074058	156698854	0	9.55%	7.99%	7.94%
chr7	1553192.5	1819257	0	9.04%	5.15%	6.74%
chr13	36110868	36738072	0	9.38%	6.60%	9.01%
chr3	116996975	117836331	0	11.38%	12.70%	11.22%
chr15	96715164	97345374	0	9.69%	4.37%	3.56%
chr2	6086404	6920924	0	8.93%	7.21%	11.22%
chr8	34182528	35417823	0	10.59%	9.46%	12.24%
chr2	79312444	79587308	0	11.71%	11.98%	12.97%
chr9	138289504	138619497	0	8.72%	6.14%	7.76%
chr16	19971547	20493661	0	9.40%	6.32%	8.86%
chr8	23830907	24507197	0	9.57%	5.73%	8.15%
chr5	155133584	155877178	0	8.80%	7.04%	9.18%
chr7	18043678	18832603.5	0	9.40%	3.32%	7.53%
chr15	95847523	96538912	0	8.64%	3.21%	9.17%
chr3	24961439	25635650	0	9.53%	12.37%	7.88%
chr16	77797358	78024002	0	9.73%	4.96%	7.55%
chr5	9623366	10150122.5	0	10.16%	8.08%	8.56%
chr1	76251948	76723043	0	9.90%	7.18%	8.77%
chr3	118144311	118792317	0	11.31%	9.02%	9.93%
chr11	40136810	40359642	0	14.82%	14.61%	11.88%
chr5	127744238	128150396	0	11.02%	5.40%	8.78%
chr4	72608149	73020263	0	10.85%	6.34%	10.79%
chr15	87410878	88053404	0	11.68%	8.60%	11.32%
chr6	153155784	153552343.5	0	10.45%	5.83%	6.77%
chr6	133908606	134175949	0	9.77%	7.41%	8.28%

Chr	Start	End	FWER	OE vs. YP	SCC vs. Normal Samples	WGBS Diff
chr6	133480830	133788156	0	11.91%	7.13%	9.17%
chr3	33910459	34429403	0	10.61%	6.54%	8.54%
chr5	174543104	174862224	0	11.63%	6.30%	7.69%
chr6	126984891	127535312.5	0	11.74%	2.79%	11.27%
chr2	124440055	125086435	0	11.78%	6.03%	10.15%
chr1	159408079	159685185	0	11.85%	12.40%	10.86%
chr10	69146234	69456351.5	0	12.37%	8.72%	9.21%
chr8	54943820	55356479	0	11.36%	6.21%	5.55%
chr15	60233780	60455085	0	-10.33%	-5.56%	0.05%
chr2	135290945	135625209	0	11.05%	4.71%	7.76%
chr7	11730278	12376625	0	10.70%	7.08%	9.71%
chr7	157845265	158155974	0.001	7.57%	4.94%	4.60%
chr4	187664780	187869873	0.001	11.13%	4.75%	4.48%
chr8	96694077	97249283	0.001	9.77%	2.90%	7.28%
chr20	29844562	30126382	0.001	9.76%	4.49%	6.56%
chr4	176339367	177163167	0.001	10.62%	4.34%	8.58%
chr12	62404795	62685785	0.001	15.65%	7.25%	6.19%
chr8	105508496	105962177	0.001	11.84%	7.39%	4.17%
chr7	61968148	62198343	0.001	8.92%	7.02%	10.37%
chr22	48792702	49080471	0.001	11.45%	8.50%	8.59%
chr19	54947714	55459006	0.002	6.61%	5.10%	7.73%
chr11	59836817	60553755	0.002	7.04%	7.27%	10.37%
chr11	124120737	124487316	0.002	10.22%	9.28%	9.80%
chr11	56264709.5	56643145	0.002	8.47%	7.59%	8.62%
chr4	95762700.5	96761411	0.002	8.86%	4.23%	8.64%
chr5	40905255	41418757	0.002	8.65%	9.22%	11.88%
chr3	29623882	30140788	0.002	9.42%	2.51%	8.88%
chr4	92879355	93443138	0.002	15.66%	10.38%	8.99%
chr12	81077928	81349606	0.002	10.00%	7.54%	6.33%
chr16	85873323	86431983	0.003	6.79%	5.27%	6.86%
chr2	79823717	80557293	0.003	8.46%	6.08%	11.13%
chr16	1093274	1331552	0.003	8.73%	8.12%	9.42%
chr13	47316638	48046005	0.003	7.99%	6.33%	8.68%
chr7	90149438	90682867	0.003	8.62%	2.94%	9.43%
chr11	91536479	92150122	0.003	8.68%	8.47%	10.45%
chr2	5201639	5749552	0.003	8.33%	7.12%	9.96%
chr7	141674839	141999568	0.003	8.26%	7.25%	10.37%
chr11	134479385	134911302	0.003	9.43%	5.68%	3.30%
chr6	45709808	46112967	0.003	10.55%	5.02%	7.89%

Chr	Start	End	FWER	OE vs. YP	SCC vs. Normal Samples	WGBS Diff
chr1	247587663.5	247921656	0.003	10.09%	6.49%	9.03%
chr6	118480809	118776527	0.003	12.38%	8.78%	9.01%
chr12	21283725	21547892	0.004	14.62%	9.35%	11.36%
chr1	101094117	101338750	0.004	9.42%	6.63%	7.97%
chr5	101925501	102211822	0.004	14.12%	2.34%	3.77%
chr11	23919652	24346161	0.004	9.32%	6.81%	12.48%
chr8	89687477	90132979	0.004	14.35%	14.21%	13.40%
chr13	76869358.5	77091333	0.004	13.02%	7.40%	10.61%
chr14	47096230	47572921	0.004	13.29%	12.31%	10.62%
chr4	18399893	19197380	0.004	12.62%	6.49%	10.21%
chr4	70808994	71346433	0.005	8.00%	10.28%	9.66%
chr3	139557385	140587268	0.005	8.22%	5.03%	8.08%
chr10	31300871	31907941	0.005	8.41%	0.54%	6.34%
chr1	237272840	238027783	0.005	7.77%	8.84%	9.93%
chr5	33563927	33901508	0.005	9.81%	6.59%	8.68%
chr13	107769911	108110276	0.005	9.32%	10.38%	10.90%
chr8	56131518	56437701.5	0.005	8.50%	5.91%	6.50%
chr3	182833979.5	183049041	0.005	10.29%	3.70%	1.13%
chr2	2176774	2402494.5	0.005	9.86%	10.16%	8.50%
chr12	20556313	20848553	0.005	10.69%	7.18%	10.49%
chr8	134594669	135056370	0.006	8.30%	4.11%	6.24%
chr7	24007534	24517414	0.006	8.16%	0.86%	6.65%
chr4	177250214	177654753	0.006	8.73%	4.48%	9.43%
chr5	136112615	136530712	0.006	8.93%	9.52%	6.30%
chr3	172635656	173001443	0.006	9.97%	11.26%	11.58%
chr12	28570321	29021552	0.006	11.20%	6.35%	3.17%
chr16	58885137	59363443	0.006	10.12%	10.08%	11.58%
chr6	29080188	29497165	0.007	7.57%	7.18%	10.69%
chr11	6865733.5	7369177	0.007	7.62%	4.44%	4.19%
chr3	115079279	115524110.5	0.007	9.01%	6.58%	7.38%
chr6	114466051	114898257	0.007	9.14%	9.66%	5.42%
chr13	94764647	95133339	0.007	9.21%	8.83%	6.57%
chr22	44220872	44682494	0.007	9.59%	0.93%	3.46%
chr7	17138805	17460742	0.007	11.96%	0.46%	1.58%
chr5	87018149	87246732	0.007	12.79%	8.28%	12.03%
chr3	2696551	3184210	0.008	7.90%	6.62%	8.53%
chr7	93540134	93919388	0.008	8.66%	6.15%	9.02%
chr3	108474544	108953233	0.008	8.58%	10.56%	9.34%
chr3	167172692	167659189	0.008	9.16%	7.24%	5.18%

Chr	Start	End	FWER	OE vs. YP	SCC vs. Normal Samples	WGBS Diff
chr1	240450031	240723218	0.008	10.19%	8.31%	9.49%
chr16	47733368	48157560	0.009	8.86%	7.12%	7.29%
chr2	167115691	167395901	0.009	8.60%	6.69%	9.26%
chr8	28984472	29316633	0.009	8.17%	1.85%	1.28%
chr11	30488071	30771240	0.009	10.79%	3.88%	7.84%
chr11	1043901.5	1307160	0.01	7.53%	7.56%	7.15%
chr4	113776216	114288822	0.01	8.16%	9.19%	9.39%
chr15	26860777	27300396	0.01	7.65%	5.70%	6.52%
chr13	36927379	37422671	0.01	8.29%	4.81%	7.60%
chr4	127771714	128125758	0.01	9.63%	7.32%	11.62%
chr8	69216882	69539427	0.01	12.37%	6.39%	8.40%
chr5	30390070	31048607	0.011	10.23%	11.58%	11.88%
chr11	130271835	130642799	0.012	6.66%	3.61%	6.42%
chr14	62343481	62550952.5	0.012	8.88%	7.81%	9.32%
chr4	20006267	20529826.5	0.012	10.00%	7.18%	7.34%
chr10	129910440.5	130268807.5	0.013	8.05%	6.17%	4.78%
chr20	31482267	31830434	0.014	7.68%	6.75%	5.62%
chr1	4000362.5	4275141	0.014	6.38%	6.84%	8.23%
chr10	10099487	10631670	0.014	8.64%	9.96%	10.56%
chr5	113512006	113806337	0.014	9.03%	7.58%	8.26%
chr2	50623873	51074910	0.014	8.76%	10.26%	7.85%
chr11	121593844.5	121835505	0.014	7.66%	-1.99%	8.90%
chr3	131538721	131855983	0.014	12.65%	5.81%	9.23%
chr12	55524042	55945613	0.015	7.78%	8.59%	9.84%
chr7	155613378	155828267	0.015	8.90%	8.27%	7.82%
chr11	28399958	28857582	0.015	7.80%	5.29%	4.62%
chr3	116428355.5	116835240	0.015	8.33%	8.41%	10.77%
chr2	48920986	49271470	0.015	8.24%	8.44%	6.14%
chr22	44963721	45330569	0.015	8.22%	3.82%	5.82%
chr13	67676107	68261695	0.015	10.41%	7.66%	10.62%
chr8	35501934	35762637	0.015	11.87%	8.09%	12.91%
chr3	188618380	188868983	0.015	9.33%	0.42%	1.18%
chr3	161235909	161577612	0.015	9.55%	8.01%	10.49%
chr16	77098093	77574072	0.016	7.89%	4.66%	8.54%
chr1	157509149	157938690	0.017	7.86%	10.59%	9.46%
chr5	166419955	166938213	0.017	8.62%	5.69%	8.31%
chr14	79760047	79968686	0.017	8.56%	6.52%	3.73%
chr10	52566320	53071348	0.018	7.68%	4.66%	6.54%
chr1	239287722	240029733	0.018	7.73%	4.59%	11.25%

Chr	Start	End	FWER	OE vs. YP	SCC vs. Normal Samples	WGBS Diff
chr1	77477593	77879031	0.018	8.57%	7.24%	4.73%
chr6	72178744	72455356	0.018	11.25%	3.19%	8.47%
chr5	88081844	88432540	0.018	10.49%	5.96%	4.02%
chr5	156900363	157107153	0.019	8.43%	1.06%	4.64%
chr3	19026539	19521725	0.02	7.98%	3.77%	11.03%
chr8	105244257	105453719	0.02	9.18%	3.90%	6.49%
chr2	143987421	144303521	0.02	9.62%	0.79%	9.27%
chr2	11024896	11245460.5	0.021	8.94%	2.01%	4.14%
chr12	11508049	11767243	0.021	8.69%	5.00%	6.55%
chr8	131705773	132404163	0.021	9.48%	6.88%	10.38%
chr8	15873055	16240240	0.021	10.25%	9.20%	10.47%
chr8	40319233	40811496	0.022	7.22%	8.26%	6.21%
chr12	15501337	15983280	0.022	6.80%	3.35%	4.56%
chr7	119831208	120689323	0.022	7.60%	4.69%	8.30%
chr7	43177046	43435880	0.022	8.14%	6.16%	4.04%
chr3	148360785	148592849	0.022	8.30%	9.82%	6.08%
chr10	44426803	44910229	0.023	7.20%	7.65%	5.45%
chr16	71008391	71221950	0.023	7.88%	2.26%	7.89%
chr4	173470589	173905761	0.023	10.20%	10.16%	10.31%
chr14	19553612	19829406	0.023	8.95%	4.92%	7.76%
chr2	225965431	226439867	0.023	11.55%	7.47%	6.25%
chr2	1507763	1732313	0.024	7.09%	1.98%	6.17%
chr1	198466913	198850158	0.024	7.53%	7.77%	6.94%
chr3	135631071	135957815	0.024	7.95%	3.06%	0.94%
chr6	70762024	71136165	0.024	10.22%	7.62%	6.83%
chr17	77173076	77702964	0.025	5.38%	4.39%	7.11%
chr8	108238734	108511437	0.026	7.97%	5.40%	8.38%
chrX	140673351	140997122	0.026	7.26%	9.20%	5.10%
chr7	3041094.5	3313269	0.027	7.08%	3.63%	4.50%
chr19	51883242	52114809	0.027	7.31%	4.58%	4.43%
chr5	146812118	147052436	0.027	7.03%	4.60%	4.07%
chr13	103702677	104168550	0.027	8.65%	7.31%	7.70%
chr17	3056402	3379425	0.028	7.06%	5.31%	5.84%
chr2	100498600	100926650	0.031	6.69%	6.10%	7.17%
chr6	134609828	134963096	0.031	7.26%	8.08%	1.14%
chr1	107224529	107886561	0.031	8.80%	4.60%	4.03%
chr7	121895636	122636635	0.032	6.98%	2.36%	6.85%
chr6	24704389	24950870	0.032	7.57%	4.10%	3.88%
chr2	236139863	236417406	0.032	7.06%	1.79%	4.28%

Chr	Start	End	FWER	OE vs. YP	SCC vs. Normal Samples	WGBS Diff
chr3	11939930.5	12269362	0.032	7.83%	4.01%	5.25%
chr4	69207281	69681824	0.032	8.33%	5.70%	8.67%
chr13	98507990	98968161	0.035	6.00%	3.68%	6.40%
chr6	153938192	154429895	0.035	9.57%	4.25%	11.65%
chr1	196746875	197037618	0.035	8.94%	7.14%	7.93%
chrX	118750273	119030658	0.035	5.27%	0.40%	0.93%
chr16	86917536	87119252	0.036	6.02%	2.98%	6.07%
chr11	89232216	89735737	0.037	7.31%	7.45%	10.57%
chr13	108434882.5	108953856.5	0.037	7.45%	1.77%	9.28%
chr8	122598941	122863171	0.037	9.13%	-0.05%	3.61%
chr1	238341559	238811442	0.037	12.33%	2.88%	6.14%
chr11	19598794	19860035	0.038	7.69%	5.35%	3.12%
chr4	5865654	6196445	0.038	6.58%	3.44%	6.05%
chr2	179544947	179974586	0.039	6.54%	7.82%	6.98%
chr15	59903988.5	60110153	0.039	9.61%	2.85%	2.94%
chr7	37873301.5	38269241	0.04	7.26%	3.62%	9.01%
chr3	173574358	174048079	0.04	8.23%	10.16%	8.51%
chr17	54230707.5	54509207	0.04	9.23%	3.29%	9.00%
chr13	43119835	43395572	0.041	7.65%	1.64%	8.49%
chr8	76147299	76477653	0.041	10.14%	10.07%	9.00%
chr5	169482268	169780607.5	0.042	7.41%	8.80%	8.32%
chr5	82109937	82395650	0.043	7.12%	4.55%	4.73%
chr12	72071095	72477363	0.044	6.96%	4.43%	4.71%
chr5	64372500	64666244	0.044	7.38%	4.70%	8.58%
chr13	112800009.5	113049747	0.045	5.77%	3.73%	6.64%
chr3	16738417	17082105	0.046	6.66%	6.24%	8.16%
chr17	76704594	76936927	0.049	5.95%	0.32%	2.84%
chr15	38641673	39145977	0.049	6.25%	3.80%	3.24%
chr14	20247808	20692622	0.049	7.20%	7.17%	8.82%
chr2	104961016	105200688	0.049	7.68%	6.60%	9.45%

Table 1.3: YE-YP Blocks

Chr	Start	End	FWER	Width (bp)
chr15	96797664	97250927	0	453263
chr5	101925501	102193800	0	268299
chr6	126984891	127484100	0.001	499209
chr6	134570471	134963096	0.002	392625
chr6	1836850	2134144	0.002	297294
chr13	100376969	100578708	0.005	201739
chr2	177510485	177872615	0.016	362130
chr15	47618795	47967175	0.019	348380
chr5	166419955	166620727	0.022	200772
chr7	17138805	17424006	0.037	285201
chr1	50907107	51168010	0.047	260903
chr8	128191730	128403369	0.048	211639

Table 1.4:OE-YE Blocks

Chr	Start	End	FWER	Width
chr11	4791012	5474774.5	0	683762.5
chr11	55541269	56143818	0	602549
chr1	158223723.5	158799979.5	0	576256
chr14	101302507	101595517.5	0.001	293010.5
chr1	2615690	3143745	0.001	528055
chr3	164621774	164915196	0.001	293422
chr14	106053344	106410681	0.002	357337
chr13	94461196	95022023	0.003	560827
chr2	79312444	79515161.5	0.003	202717.5
chr5	155379667	155956646	0.004	576979
chr5	156110742	156642345	0.005	531603
chr16	19971547	20493661	0.005	522114
chr1	248307095	248637451	0.005	330356
chr11	40136810	40359642	0.005	222832
chr4	16532808	16795757	0.006	262949
chr7	136563350	136953057	0.008	389707
chr9	138362412.5	138586311	0.01	223898.5
chr3	25101907	25493807	0.01	391900
chr6	29232436	29443423	0.016	210987
chr7	61968148	62198343	0.016	230195
chr11	124172447	124413717	0.017	241270
chr8	34843233	35297622	0.018	454389
chr5	127744238	128034864	0.021	290626
chr11	91717555	92116826	0.023	399271
chr12	97924358	98178057	0.023	253699
chr3	116996975	117329281	0.024	332306
chr8	89687477	90132979	0.025	445502
chr7	93697091	93984889	0.031	287798
chr1	157567720	157811604	0.031	243884
chr15	87436107	87892167	0.037	456060
chr6	133529701	133757084	0.048	227383
chr6	153224221	153552343.5	0.05	328122.5
chr7	11676020.5	12376625	0.05	700604.5

Table 1.5: Block Finding Summary

Epidermis	Number 450k	Total Size 450k	Mean Change 450k	Total Size WGBS	Mean Change WGBS
O-exp vs Y-pro (Sun + Age)	224	99 Mb	9.20%	670 Mb	9.00%
Y-exp vs Y-pro (Sun/Site)	12	3.8 Mb	9.30%	0.9 Mb	7.30%
O-exp vs O-pro	239	100 Mb	8.60%	1.5 Mb	7.50%
O-pro vs Y-pro (Age)	0	-	-	6.2 Mb	7.10%
O-exp vs Y-exp	33	12.7 Mb	10.00%	8.5 Mb	8.00%
Dermis	**Number**	**Total Width**	**Mean Difference**		
O-exp vs Y-pro	0	-	-	NA	NA
Y-exp vs Y-pro	0	-	-	NA	NA
O-pro vs Y-pro	0	-	-	NA	NA
O-exp vs Y-exp	0	-	-	NA	NA

Table 1.6: WGBS Alignment Summary

Sample	# paired-ends reads sequenced	# ends sequenced (reads x2)	# ends aligned	alignment rate
Older, Exposed 1	147552493	295104986	244101375	0.83
Older, Exposed 2	153586791	307173582	249181172	0.81
Older, Exposed 3	147428801	294857602	243467309	0.83
Older, Protected 1	174946497	349892994	285964233	0.82
Older, Protected 2	130095651	260191302	212045113	0.81
Older, Protected 3	152377633	304755266	252208249	0.83
Younger, Exposed 1	157543786	315087572	257584634	0.82
Younger, Exposed 2	170708509	341417018	284253113	0.83
Younger, Exposed 3	148069990	296139980	245818274	0.83
Younger, Protected 1	173772291	347544582	281444473	0.81
Younger, Protected 2	177973878	355947756	290919238	0.82
Younger, Protected 3	163700948	327401896	265120593	0.81

Table 1.7: WGBS BSmooth Summary

Sample	Covered CpGs	coverage	depth (X)	Methylated CpGs	Methylated CpG %
Older, Exposed 1	24609491	87.21%	5.68	21418192	87.03%
Older, Exposed 2	24869681	88.14%	5.83	21430311	86.17%
Older, Exposed 3	24865651	88.12%	5.71	21510892	86.51%
Older, Protected 1	24945707	88.41%	6.48	21933497	87.92%
Older, Protected 2	24387124	86.43%	4.97	21084564	86.46%
Older, Protected 3	25055896	88.80%	5.96	21893277	87.38%
Younger, Protected 3	24853692	88.08%	6.07	21846635	87.90%
Younger, Exposed 1	24948670	88.42%	6.12	21862022	87.63%
Younger, Exposed 2	25047251	88.77%	6.55	21967490	87.70%
Younger, Exposed 3	24803372	87.90%	5.71	21719843	87.57%
Younger, Protected 1	24787671	87.85%	6.44	21810045	87.99%
Younger, Protected 2	25311893	89.70%	6.77	22221374	87.79%

Table 1.8:GSEA Results

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
ZHANG_TLX_TARGETS_36HR_DN	152	-0.74	-3.16	0	0	0
PUJANA_XPRSS_INT_NETWORK	150	-0.67	-2.92	0	0	0
ZHANG_TLX_TARGETS_UP	73	-0.75	-2.88	0	0	0
SOTIRIOU_BREAST_CANCER_GRADE_1_VS_3_UP	130	-0.67	-2.82	0	0	0
WINNEPENNINCKX_MELANOMA_METASTASIS_UP	131	-0.66	-2.75	0	0	0
CHANG_CYCLING_GENES	127	-0.65	-2.75	0	0	0
DUTERTRE_ESTRADIOL_RESPONSE_24HR_UP	289	-0.58	-2.72	0	0	0
PUJANA_BRCA_CENTERED_NETWORK	106	-0.66	-2.69	0	0	0
PYEON_CANCER_HEAD_AND_NECK_VS_CERVICAL_UP	152	-0.63	-2.68	0	0	0
KANG_DOXORUBICIN_RESISTANCE_UP	49	-0.74	-2.68	0	0	0
ZHANG_TLX_TARGETS_60HR_DN	234	-0.59	-2.68	0	0	0
FERREIRA_EWINGS_SARCOMA_UNSTABLE_VS_STABLE_UP	140	-0.62	-2.65	0	0	0
ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER	125	-0.64	-2.64	0	0	0
ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_24HR	108	-0.65	-2.63	0	0	0
SHEN_SMARCA2_TARGETS_UP	348	-0.55	-2.62	0	0	0
ZHANG_BREAST_CANCER_PROGENITORS_UP	352	-0.55	-2.61	0	0	0
MOLENAAR_TARGETS_OF_CCND1_AND_CDK4_DN	47	-0.74	-2.61	0	0	0
MILI_PSEUDOPODIA_HAPTOTAXIS_UP	416	-0.53	-2.56	0	0	0
ZHOU_CELL_CYCLE_GENES_IN_IR_RESPO	72	-0.67	-2.55	0	0	0

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
NSE_6HR						
FARMER_BREAST_CANCER_CLUSTER_2	31	-0.79	-2.55	0	0	0
ISHIDA_E2F_TARGETS	47	-0.73	-2.55	0	0	0
MORI_IMMATURE_B_LYMPHOCYTE_DN	82	-0.65	-2.53	0	0	0
SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_A6	391	-0.52	-2.52	0	0	0
BIDUS_METASTASIS_UP	167	-0.58	-2.52	0	0	0
CROONQUIST_NRAS_SIGNALING_DN	69	-0.66	-2.51	0	0	0
KONG_E2F3_TARGETS	84	-0.64	-2.50	0	0	0
TANG_SENESCENCE_TP53_TARGETS_DN	52	-0.69	-2.50	0	0	0
BURTON_ADIPOGENESIS_3	92	-0.63	-2.50	0	0	0
PUJANA_BREAST_CANCER_WITH_BRCA1_MUTATED_UP	49	-0.70	-2.48	0	0	0
WHITEFORD_PEDIATRIC_CANCER_MARKERS	104	-0.60	-2.46	0	0	0
KAUFFMANN_MELANOMA_RELAPSE_UP	57	-0.67	-2.46	0	0	0
GABRIELY_MIR21_TARGETS	250	-0.53	-2.45	0	0	0
MORI_LARGE_PRE_BI_LYMPHOCYTE_UP	79	-0.62	-2.45	0	0	0
BENPORATH_PROLIFERATION	127	-0.59	-2.45	0	0	0
KOBAYASHI_EGFR_SIGNALING_24HR_DN	221	-0.54	-2.45	0	0	0
PUJANA_BRCA2_PCC_NETWORK	378	-0.51	-2.44	0	0	0
MOREAUX_MULTIPLE_MYELOMA_BY_TACI_DN	141	-0.58	-2.44	0	0	0
LEE_EARLY_T_LYMPHOCYTE_UP	89	-0.61	-2.42	0	0	0
KAMMINGA_EZH2_TARGETS	39	-0.71	-2.42	0	0	0
CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION_UP	152	-0.56	-2.41	0	0	0

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
HOFFMANN_LARGE_TO_SMALL_PRE_BII_LYMPHOCYTE_UP	145	-0.56	-2.39	0	0	0
CROONQUIST_IL6_DEPRIVATION_DN	90	-0.60	-2.38	0	0	0
GARY_CD5_TARGETS_DN	370	-0.49	-2.36	0	0	0
SONG_TARGETS_OF_IE86_CMV_PROTEIN	57	-0.64	-2.36	0	0	0
WHITFIELD_CELL_CYCLE_LITERATURE	41	-0.69	-2.36	0	0	0
MARKEY_RB1_ACUTE_LOF_DN	201	-0.53	-2.36	0	0	0
REICHERT_MITOSIS_LIN9_TARGETS	26	-0.75	-2.35	0	0	0
PYEON_HPV_POSITIVE_TUMORS_UP	69	-0.62	-2.35	0	0	0
EGUCHI_CELL_CYCLE_RB1_TARGETS	22	-0.79	-2.34	0	0	0
ZHAN_MULTIPLE_MYELOMA_PR_UP	36	-0.70	-2.34	0	0	0
REACTOME_DNA_REPLICATION	166	-0.53	-2.32	0	0	0
WHITFIELD_CELL_CYCLE_G1_S	117	-0.56	-2.30	0	0	0
GREENBAUM_E2A_TARGETS_UP	33	-0.72	-2.30	0	0	0
REN_BOUND_BY_E2F	56	-0.64	-2.29	0	0	0
DAZARD_RESPONSE_TO_UV_NHEK_DN	271	-0.49	-2.27	0	0	0
IWANAGA_E2F1_TARGETS_INDUCED_BY_SERUM	25	-0.74	-2.27	0	1.34E-05	0.001
ODONNELL_TFRC_TARGETS_DN	117	-0.54	-2.27	0	1.32E-05	0.001
REACTOME_MITOTIC_PROMETAPHASE	74	-0.58	-2.26	0	1.30E-05	0.001
MITSIADES_RESPONSE_TO_APLIDIN_DN	215	-0.50	-2.26	0	1.27E-05	0.001
REACTOME_CELL_CYCLE_MITOTIC	271	-0.49	-2.26	0	1.25E-05	0.001
VANTVEER_BREAST_CANCER_METASTASIS_DN	103	-0.56	-2.25	0	1.23E-05	0.001
GRAHAM_NORMAL_QUIESCENT_VS_NORMAL_DIVIDING_DN	78	-0.58	-2.25	0	1.21E-05	0.001

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
WU_APOPTOSIS_BY_CDKN1A_VIA_TP53	47	-0.65	-2.24	0	1.19E-05	0.001
RHODES_UNDIFFERENTIATED_CANCER	63	-0.59	-2.24	0	1.17E-05	0.001
KAUFFMANN_DNA_REPAIR_GENES	213	-0.49	-2.23	0	2.30E-05	0.002
FUJII_YBX1_TARGETS_DN	186	-0.50	-2.23	0	2.26E-05	0.002
MISSIAGLIA_REGULATED_BY_METHYLATION_DN	107	-0.54	-2.22	0	3.37E-05	0.003
SCIBETTA_KDM5B_TARGETS_DN	70	-0.58	-2.22	0	3.32E-05	0.003
REACTOME_MITOTIC_M_M_G1_PHASES	147	-0.52	-2.22	0	4.38E-05	0.004
MOREAUX_B_LYMPHOCYTE_MATURATION_BY_TACI_DN	58	-0.61	-2.22	0	5.37E-05	0.005
DAZARD_UV_RESPONSE_CLUSTER_G6	126	-0.53	-2.21	0	6.40E-05	0.006
REACTOME_GENERIC_TRANSCRIPTION_PATHWAY	305	-0.47	-2.20	0	7.35E-05	0.007
VERNELL_RETINOBLASTOMA_PATHWAY_UP	63	-0.59	-2.19	0	7.25E-05	0.007
AMUNDSON_GAMMA_RADIATION_RESPONSE	36	-0.66	-2.19	0	8.18E-05	0.008
GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_UP	460	-0.45	-2.19	0	8.07E-05	0.008
PID_FANCONI_PATHWAY	42	-0.62	-2.17	0	1.39E-04	0.014
LI_WILMS_TUMOR_ANAPLASTIC_UP	16	-0.78	-2.16	0	1.66E-04	0.016
DACOSTA_UV_RESPONSE_VIA_ERCC3_COMMON_DN	411	-0.45	-2.16	0	1.64E-04	0.016
GEORGES_CELL_CYCLE_MIR192_TARGETS	59	-0.59	-2.16	0	1.62E-04	0.016
FEVR_CTNNB1_TARGETS_DN	455	-0.44	-2.16	0	1.70E-04	0.017
REACTOME_CELL_CYCLE	345	-0.45	-2.15	0	2.05E-04	0.021
REACTOME_G2_M_CHECKPOINTS	32	-0.65	-2.13	0	2.49E-04	0.026
CHEMNITZ_RESPONS	118	-0.51	-2.12	0	3.19E-	0.034

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
E_TO_PROSTAGLANDIN_E2_UP					04	
GRAHAM_CML_DIVIDING_VS_NORMAL_QUIESCENT_UP	164	-0.49	-2.11	0	3.33E-04	0.036
ODONNELL_TARGETS_OF_MYC_AND_TFRC_DN	37	-0.62	-2.11	0	3.73E-04	0.041
THUM_SYSTOLIC_HEART_FAILURE_DN	196	-0.47	-2.11	0	3.78E-04	0.042
FINETTI_BREAST_CANCER_KINOME_RED	15	-0.79	-2.11	0	4.00E-04	0.045
GAZDA_DIAMOND_BLACKFAN_ANEMIA_PROGENITOR_DN	54	-0.57	-2.10	0	4.30E-04	0.049

Table 1.9: OE-YP Small DMRs

Chr	Start	End	FWER	Mean Difference in Methylation- OE vs YP Samples
chr7	27196759	27198896	0	37.68%
chr5	134362967	134364717	0	37.19%
chr16	85320035	85320882	0	-41.31%
chr17	4648566	4649262	0	43.14%
chr7	27199726	27199956	0.001	52.39%
chr7	27191097	27192656	0.003	33.42%
chr7	134832221	134833166	0.003	-34.00%
chr4	54959419	54960259	0.003	34.65%
chr10	3158688	3159145	0.003	49.57%
chr13	25085301	25085669	0.003	-51.81%
chr7	27183861	27185512	0.004	31.25%
chr4	1352787	1353919	0.004	-34.19%
chr12	2944163	2944493	0.004	34.29%
chr15	102009795	102010195	0.004	-34.01%
chr1	243584668	243584861	0.004	-50.62%
chr4	187762256	187762379	0.005	-49.33%
chr6	32819858	32820249	0.006	-32.92%
chr15	37389384	37390326	0.006	33.37%
chr8	141577061	141577526	0.006	33.61%
chr6	161860203	161860561	0.006	-47.09%
chr7	2076633	2076633	0.007	-63.63%
chr2	172951828	172953925	0.01	30.58%
chr2	127978490	127978579	0.011	-46.61%
chr7	2654420	2654420	0.011	-61.49%
chr6	33288561	33289208	0.013	-26.61%
chr22	51016501	51017019	0.013	30.94%
chr11	31818791	31819678	0.013	31.58%
chr7	862427	862831	0.013	-31.97%
chr17	80255419	80255510	0.013	-45.39%
chr5	80529067	80529340	0.014	31.37%
chr8	116679763	116680127	0.014	-40.33%
chr17	8129997	8130356	0.014	44.70%
chr8	74903761	74903801	0.014	-44.52%
chr7	19149989	19150174	0.014	-44.46%
chr17	73805953	73805997	0.015	-44.31%
chr12	115134201	115135333	0.016	-25.75%
chr7	27154562	27155358	0.016	29.38%
chr19	16830613	16830859	0.016	-38.47%

Chr	Start	End	FWER	Mean Difference in Methylation- OE vs YP Samples
chr2	223917577	223918000	0.016	39.13%
chr5	137803102	137803362	0.016	-38.93%
chr3	160121821	160122503	0.016	-38.76%
chr6	121767874	121767882	0.017	-44.15%
chr17	40715222	40715281	0.019	-38.09%
chr15	63220652	63220652	0.019	-56.50%
chr12	115103960	115105710	0.02	-28.64%
chr7	157866922	157867812	0.02	-30.15%
chr19	49340489	49340765	0.02	38.16%
chr15	31774773	31775049	0.02	-42.97%
chr7	130626376	130626559	0.021	-42.53%
chr8	12668973	12669347	0.021	-41.56%
chr1	94560697	94560851	0.022	-43.18%
chr6	31698218	31698226	0.025	40.98%
chr16	66612955	66613355	0.028	28.50%
chr11	31827084	31828040	0.028	28.84%
chr11	31820441	31822526	0.029	26.96%
chr2	135594780	135594780	0.029	-54.29%
chr11	268923	269468	0.03	-35.98%
chr2	172945144	172947075	0.031	28.24%
chr17	76220608	76220955	0.031	29.18%
chr11	65546988	65547172	0.031	-36.30%
chr2	172956823	172958358	0.032	28.65%
chr6	2891973	2892152	0.032	-35.76%
chr17	80194576	80195737	0.032	-29.04%
chr11	31825756	31825969	0.033	35.84%
chr12	133156510	133156565	0.033	39.62%
chr12	115129011	115129801	0.034	39.39%
chr6	169051407	169051431	0.035	39.13%
chr3	188697399	188697399	0.035	-53.60%
chr2	100279893	100279893	0.035	-53.50%
chr11	70001754	70002283	0.036	-34.74%
chr10	6117324	6117324	0.036	-53.27%
chr7	140227195	140227335	0.037	-38.47%
chr3	187086147	187086530	0.038	-34.96%
chr2	54086854	54087517	0.039	26.19%
chr7	101723207	101723947	0.039	-27.98%
chr16	1583810	1584118	0.039	27.92%
chr7	110838461	110838461	0.04	-52.65%

Chr	Start	End	FWER	Mean Difference in Methylation- OE vs YP Samples
chr7	19152017	19152335	0.043	-34.44%
chr6	28921103	28922226	0.044	25.85%
chr5	54518667	54519307	0.044	27.57%
chr7	155252201	155252944	0.045	33.66%
chr6	29943188	29943677	0.047	25.80%
chr15	37394754	37395171	0.049	33.36%
chr7	157179773	157179830	0.05	-37.23%
chr17	77134303	77134303	0.05	-51.66%
chr19	10370550	10370550	0.05	-51.61%
chr8	69879951	69879951	0.05	-51.50%
chr12	2046045	2046230	0.052	-39.36%
chr14	101459547	101459591	0.052	-39.25%
chr2	223916502	223916861	0.053	26.98%
chr15	81666393	81666671	0.053	-37.15%
chr10	8090846	8090924	0.053	-39.12%
chr6	118672731	118672731	0.053	-51.41%
chr16	1231407	1232363	0.054	-26.88%
chr14	104184551	104184575	0.054	39.04%
chr11	65816463	65816819	0.055	26.85%
chr11	31845148	31845639	0.055	33.63%
chr4	54554571	54554571	0.055	-51.02%
chr14	24640947	24642219	0.056	24.12%
chr7	155255122	155255791	0.056	32.90%
chr15	32933704	32934185	0.056	26.77%
chr6	32805398	32805692	0.056	-26.74%
chr1	245812674	245812674	0.057	50.75%
chr6	31650735	31651158	0.059	22.99%
chr7	27140797	27142100	0.059	26.13%
chr11	31824973	31825226	0.059	33.53%
chr11	19372012	19372014	0.059	38.31%
chr7	3186766	3186788	0.059	-38.31%
chr12	115137528	115138766	0.06	-26.09%
chr16	15595889	15596215	0.062	-26.51%
chr10	820767	821718	0.062	-31.75%
chr1	233085132	233085383	0.062	-36.10%
chr11	31840628	31841980	0.063	25.85%
chr16	89734986	89735184	0.064	-37.98%
chr1	207224549	207225647	0.065	-25.86%
chr13	33780078	33780362	0.066	31.37%

Chr	Start	End	FWER	Mean Difference in Methylation- OE vs YP Samples
chr3	124860616	124860871	0.066	35.97%
chr5	101925501	101925501	0.067	-50.19%
chr11	334298	334833	0.068	32.73%
chr22	44568387	44568812	0.069	-25.09%
chr6	35479628	35479910	0.07	35.83%
chr14	52241096	52241174	0.07	-37.91%
chr7	156336589	156337209	0.071	26.01%
chr16	3998752	3998816	0.072	35.71%
chr1	29587088	29587255	0.072	35.66%
chr12	12224246	12224457	0.072	-37.76%
chr1	88421825	88421825	0.072	-49.82%
chr21	44486205	44486452	0.073	-32.25%
chr3	194407860	194408965	0.075	25.08%
chr22	46457384	46457998	0.076	-35.39%
chr2	131720820	131721768	0.079	30.57%
chr12	97300410	97300429	0.079	-35.22%
chr7	155262490	155263349	0.079	35.21%
chr6	52171973	52172083	0.079	35.19%
chr11	70303464	70303534	0.079	-37.42%
chr16	88458618	88458669	0.079	-37.37%
chr3	147126703	147127012	0.08	25.72%
chr2	74875227	74875548	0.081	24.49%
chr1	228248013	228248785	0.081	31.80%
chr4	141207794	141207852	0.081	-35.12%
chr16	1145964	1146545	0.082	-25.55%
chr21	45581541	45581710	0.083	-35.02%
chr5	110427072	110427112	0.083	-37.23%
chr4	26865310	26865310	0.084	-48.89%
chr22	46455670	46455670	0.084	-48.83%
chr21	46378243	46379089	0.086	30.27%
chr11	31824262	31824327	0.086	37.14%
chr1	68084465	68084465	0.086	-48.61%
chr12	11709115	11709115	0.087	-48.54%
chr5	155956646	155956646	0.087	48.51%
chr14	106187192	106187192	0.087	-48.50%
chr12	72875216	72875216	0.087	48.49%
chr2	119602212	119603489	0.088	22.85%
chr8	58055024	58056113	0.088	25.29%
chr12	115131223	115132015	0.088	25.27%

Chr	Start	End	FWER	Mean Difference in Methylation- OE vs YP Samples
chr16	15613108	15613108	0.088	48.46%
chr16	75148330	75149501	0.091	-25.10%
chr7	5342326	5343176	0.091	-25.10%
chr7	157475548	157475737	0.091	34.64%
chr6	45784526	45784526	0.093	-48.20%
chr1	9527173	9527208	0.094	36.93%
chr13	102104707	102105440	0.097	-23.95%
chr9	139715701	139716215	0.097	34.43%
chr16	1069760	1070137	0.099	-36.79%
chr15	62853598	62853598	0.099	47.98%
chr10	4925586	4925586	0.099	-47.89%

Table 1.10: Small DMR summary

Epidermis	**Hyper**	**Hypo**
O-exp vs Y-pro (Sun + Age)	90	76
Y-exp vs Y-pro (Sun/Site)	20	10
O-exp vs O-pro	55	11
O-pro vs Y-pro (Age)	1	1
O-exp vs Y-exp	1	13
Dermis	**Hyper**	**Hypo**
O-exp vs Y-pro	6	3
Y-exp vs Y-pro	2	7
O-exp vs O-pro	2	4
O-pro vs Y-pro	1	0
O-exp vs Y-exp	0	0

Chapter 2: Cell type specific methylation change with aging in peripheral blood

Introduction

Human aging is the greatest risk factor for disease and morbidity, however the molecular basis remains poorly understood[46]. Aging is associated with functional decline and loss of homeostasis across multiple organ systems, which arises due to altered or reduced function of progenitor and somatic cells. Within the immune systems this process has been well characterized: aging is associated with increased activation of basal inflammatory pathways, but decreased ability to respond to new infections. This is attributed to both altered differentiation capacity of hematopoietic progenitor cells [77] and altered function of differentiated cells [79]. Because these changes occur primarily independently of changes to the genomic sequence and are variable among individuals, there has been significant interest in investigating how epigenetic modifications may mediate these changes.

DNA methylation is the most commonly studied epigenetic modification and plays an important role in cell identity. This modification is most frequently found in the context of CpG dinucleotides, where it can be placed by three methyltransferase enzymes: DNMT1, the maintenance methyltransferase that re-establishes the methylation pattern following DNA replication, and DNMT3a and DNMT3b, which function in de novo methylation [6]. The presence of DNA methylation in gene promoters and enhancers is inversely correlated with gene expression, likely due to alterations of local DNA structure and prevention of transcription factor binding [143]. By contrast, DNA methylation within gene bodies is linked to active transcription and regulation of splicing [144]. Maintenance of the methylome is essential for progenitor cell function, and de novo methylation is required for differentiation in some systems [34, 37].

In addition to local, gene specific changes in methylation, large scale loss of methylation has been observed with cancer and EBV-immortalization [28, 31, 145]. These changes are detectable at a global level and involve loss of methylation across large, gene-poor regions of the genome, termed “hypomethylated blocks”. The hypomethylated blocks overlap strongly with heterochromatic, lamina-associated domains. Intriguingly, similar large-scale hypomethylation in gene poor regions is observed in cells approaching replicative senescence [95]. In this work, loss of methylation was observed to occur preferentially in late-replicating regions, attributed to mislocalization of DNMT1 in late passage cells.

Both small and large-scale DNA methylation changes associated with aging have been reported in peripheral blood, however the true nature of cell intrinsic methylation changes with age remains unclear. An early study using the Illumina HumanMethylation27 Beadchip (27k) array, a promoter-biased array, showed common regions of differential methylation with age in CD4+ lymphocytes and CD14+ monocytes. In this study, age associated hypermethylation occurred preferentially in regions identified as bivalent in embryonic stem cells [85], however this study was limited by the low coverage and bias of early array technology. The release of the more genome scale Illumina HumanMethylation Beadchip450 (450k) array has facilitated more thorough observations: work by Hannum et al. analyzing a large cohort of peripheral blood samples proposed an “epigenetic clock” from which chronological age can be predicted based on methylation at 71 sites measured by 450k array[102]. Horvath refined this approach, identifying a model using 353 CpGs which is used to calculate a “methylation age” that correlates strongly with chronological age across a large variety of tissues and

cell types[89]. A recent work by Yuan et al. used the 450k array data to identify large-scale/block level hypomethylation with age in peripheral blood [146].

The observations of these studies are limited by the reliance on data from peripheral blood, which consists of multiple cell types, each of which has a distinct methylome. As individuals age, the relative proportion of cells within peripheral blood shifts, confounding analysis of methylation data [103]. While it is possible to account somewhat for these changes using linear regression [146], this correction relies on the assumption that there are no cell type specific age related changes with age. Analysis of purified blood cells has been very limited: sorted CD4+ cells from one newborn and one centenarian examined by Heyn et al. using whole genome bisulfite sequencing (WGBS) showed large scale hypomethylation with age reminiscent of the hypomethylated blocks, however the generalizability of this result is limited by the use of only one cell type and single donors in each age group[94]. Another recent analysis of 450k array data from CD14+ monocytes and CD4+ lymphocytes from older donors indicates very distinct methylation changes in these cell types with age but focuses on donors over 50 years of age and is reliant on the coverage of the 450k array[98].

In this work, we characterize cell type specific methylation age related methylation changes in CD14+ monocyte, CD4+ T lymphocytes and naïve CD4+ T lymphocytes isolated from donors ranging from 18 to 97 years of age. We use whole genome bisulfite sequencing in a discovery set of 3 younger and 3 older Caucasian women to avoid the biases of array based analysis and identify highly distinct patterns for each cell type. We validated these results using the 450k array on a larger donor set. Intriguingly, we observe that these changes occur in cell type specific enhancer regions

and, in naïve CD4+ cells, are enriched for binding sites for a regulatory factor involved in aging.

Results

Global hypomethylation with aging is cell type specific

In order to identify cell type specific methylation change with age, we collected peripheral blood from 43 healthy donors ranging from 18 to 97 years old (demographic information detailed in Table 2.1). From each individual, we isolated peripheral blood mononuclear cells (PBMCs) and used magnetic beads to isolate three types of blood cells, CD4+ T-lymphocytes, naïve CD4+ T-lymphocytes and CD14+ monocytes. Analysis of sorted cells using flow cytometry post-sorting indicated high purity—97% for CD4+ cells, 85% of naïve CD4+ cells, 95% for CD14+ cells (Figure 2.1).

In order to gain unbiased genome wide information about DNA methylation in these samples, we performed whole genome bisulfite sequencing (WGBS) on CD4+ and CD14+ cells from 3 of the youngest (<30 years) and 3 of the oldest donors (>80 years). To control for sex and race in this initial set, we selected only Caucasian female donors for the WGBS analysis (donors indicated in Table 2.1). We generated sequencing data to a depth of 5.1-6.0X and analyzed it using the BSmooth algorithm, which was designed for analyzing low-coverage WGBS data and has been demonstrated to accurately estimate methylation levels at single-base pair resolution by borrowing information from nearby CpGs [45]. After filtering reads with low quality measures, we obtained measurements for an average of 25,145591 CpGs per sample (average of 89% coverage).

Bisulfite conversion was assessed using spiked in lambda phage and ranged from 99.69-99.74% (Details in Methods, Tables 2.2).

To examine the potential for global methylation change with age, we began by looking at mean methylation in both cell types analyzed. Surprisingly, when considering all CpGs with coverage of 2 or more in each sample group, we observed a significant drop in mean methylation with age in CD4+ cells (mean of 81.7% methylation in the younger donors, 77.9% in older donors), but no change in global methylation with age in CD14+ cells (mean methylation of 81.5% in younger donors, 81.6% in older donors) (Figure 2.2A and 2.2B).

In previous studies of colon cancer, EBV transformation and cellular senescence, mean hypomethylation is explained by loss of mean methylation across large regions of low CpG density, overlapping with heterochromatic and lamina-associated domains [28, 31, 95]. To determine if such large regions of altered methylation are present in our dataset, we applied BSmooth's region finding algorithm to data smoothed over large (20 kbp) windows. Applying this method to the CD4+ cell data using the same cutoffs used to define blocks in EBV and colon cancer, we identified hypomethylated blocks spanning over 622 Mbp, with an average of 7.73% difference in methylation (example block region shown in Fig 2.2C.) We observed a strong overlap between the detected age hypomethylated blocks and the previously reported colon cancer hypomethylated blocks (OR=7.6). Consistent with this similarity, we also observed a significant overlap between the blocks hypomethylated with age and lamina associated domains previously mapped in fibroblasts (OR=3.3) [29] as well as H3K9me2 heterochromatic "LOCKS" mapped in fibroblasts (OR= 4.0) [28].

It has been previously reported that the proportion of naïve cells within the CD4+ T cell population declines with aging [147]. To determine impact this subpopulation has on the observed methylation changes, we examined methylation in the naïve CD4+ cells that were isolated from each donor. We observed higher mean methylation within the naïve CD4+ subset compared to the total CD4+ set for both young and old individuals. However, the regions identified as age-hypomethylated blocks within the CD4+ data are consistently hypomethylated with age in the naïve CD4+ subset (Fig 2.2D), indicating that the change in naïve CD4+ cell proportion alone cannot explain the observed methylation changes.

By contrast, when the same algorithm was applied to the CD14+ cell data, only 24 potential blocks, spanning 0.57 Mbp were identified. These blocks are both hypo and hyper methylated with age, consistent with the absence of mean methylation changes. Thus, while the widespread hypomethylation with age observed in CD4+ T-lymphocytes is consistent with previously reported loss of methylation in aging and occurs within regions that lose methylation in colon cancer, this change is cell type specific and not observed in all peripheral blood cell types.

Age-related differential methylation occurs in cell type specific regulatory regions

We next applied BSmooth's region finding algorithm to sequencing data smoothed over a smaller, 2 kb, window in order to identify specific small regions of larger magnitude differential methylation (DMRs). We used the cutoffs applied to identify regions differentially methylated in colon cancer. We identified 1487 regions differentially methylated with age in the CD4+ cells (example in Figure 2.3A). Distinct from the widespread hypomethylation, the majority of identified small DMRs (1012) are

hypermethylated with age. This pattern is similar to changes identified in cancer studies: widespread hypomethylation with focal hypermethylation[145]. A similar pattern was observed within the naïve CD4+ subset: 3645 DMRs were identified, 2150 of which are hypermethylated with age (example in Figure 2.3B). The larger number of regions identified may be due to the increased purity of the cell population, reducing noise. By contrast, for CD14+ cells, 934 DMRs were identified, the majority of which (762) are hypomethylated with age (example in Figure 2.3C), again indicating a distinct mechanism of epigenetic change in CD14+ cells with age. The results of age-DMR finding from all cell types are summarized in Table 2.3.

While within each cell type a large number of regions is identified as age-DMRs, only 5 of the hypomethylated age-DMRs were identified as hypomethylated in all cell types, and only 10 of the hypermethylated age-DMRs were identified as hypermethylated in all cell types (example in Figure 2.3E). It is possible that more common age-DMRs are present but are not of sufficient magnitude in all cell types to be detected using the relatively strict cutoffs used. Several of the genes nearest these DMRs are intriguing: one common age hypermethylated DMR is nearest to MIR30D, a microRNA involved in regulation of p53 expression [148], another, PIK3C2A, is involved in insulin signaling and activation of mTORC1 [149]. A complete list of common age DMRs is provided in Table 2.4.

Previous studies have reported overlap of hypermethylated aging related DMRs with both polycomb targets and hypomethylated aging DMRs with enhancers marked by H3K4me1[88]. To determine how the identified cell type specific DMRs relate to these marks, we took advantage of ChIP-seq data from CD14+ and naïve CD4+ cells obtained

as part of the Roadmap epigenome project[150]. For each cell type, we calculated the odds ratio for overlap between hypo and hyper methylated DMRs and ChIP-seq peaks for H3K27me3, H3K4me3 and H3K4me1. We also examined the overlap between each cell type's age-DMRs and ChIP-seq peaks mapped in the other cell type (Figure 2.4A and 2.4B). For both naïve CD4 cells and CD14 cells, we observed the strongest overlap between hypomethylated DMRs and H3K4me1 peaks mapped in the same cell type (OR= 10.3 and 7.2). The overlap between hypomethyated Age DMRs and H3K4me1 peaks mapped in the other cell type was much weaker (OR= 4.5 and 1.4), indicating that the overlap between hypomethylated age-DMRs and H3K4me1 has a high degree of cell type specificity.

Intriguingly, hypermethylated age-DMRs had a distinct pattern. In contrast to previous work examining only CpGs covered by the 27k array, we detect only a modest overlap between hypermethylated age-DMRs and H3K27me3 (OR=1.1 for naïve CD4+ and 2.7 for CD14+). We observe a stronger overlap between hypermethylated age-DMRs and H3K4me1, but not with the same cell type specificity as observed for hypomethylated DMRS: naïve CD4 hypermethylated age-DMRs have a similar degree of overlap with H3K4me1 mapped in naïve CD4 cells (OR=8.3) and in CD14 cells (OR=7.7), while CD14+ hypermethylated age-DMRs have a stronger overlap with H3K4me1 mapped in naïve CD4 cells (OR=5.9) than in CD14 cells (2.9), indicating age related hypermethylation may preferentially occur in putative enhancers not specific to the cell type.

Hypomethylated naïve CD4+ age-DMRs enriched for functional motifs

Given the observed overlap between age-DMRs and enhancers, we sought to identify transcription factors that may differentially bind to the identified regions. We used Haystack, an algorithm that looks for enrichment of known transcription factor binding sites in a set of genomic regions[151]. We separately analyzed hyper and hypomethylated DMRs from each of the three cell types sequenced (CD4+, naïve CD4+ and CD14+) against a background of all other DMRs identified to determine what distinguishes a region hypo or hypermethylated with age for a specific cell type.

The most enriched binding motif was identified as enriched within regions hypomethylated with age in naïve CD4+ cells. This motif corresponds to the binding sequence for the B-ATF::JUN complex (Fig 2.4C) and has a center enrichment of 4.35 within the identified regions (Fig 2.4D). This motif is found in 17% (265) of the hypomethylated naïve CD4+ age-DMRs. This finding is intriguing as B-ATF is activated downstream of IL-6, levels of which have been observed to increase with aging [152] and was seen to significantly increase with chronological age in serum samples collected from our donors (p-value<0.001 ,Fig 2.5A). B-ATF controls Th17 differentiation from naïve CD4 T-cells [153] and the levels of Th17 cells have been observed to change with aging [154]. We see no significant difference in IL-17, a cytokine produced exclusively by Th17 cells in serum from our donors (Fig 2.5B), however the basal levels are very low and age associated changes in IL-17 are seen only after stimulation[154].

Validation of cell type specific aging changes using the 450k array

Having observed cell type specific methylation changes related to aging in a small set of Caucasian women using whole genome bisulfite sequencing, we sought to

determine whether these changes are present in the larger, mixed race and mixed sex population. To do this, we analyzed genome wide methylation in CD4+, CD14+, naïve CD4+ and peripheral blood mononuclear cells (PBMCs) from all individuals sampled (43 donors, ranging from 18 years to 97 years, information in Table 2.1) using the Illumina Infinium HumanMethylation450 (450k) array.

We first examined CpGs covered by the 450k array that are located within the regions identified as differentially methylated from whole genome bisulfite sequencing. When we considered CpGs within the regions identified as age related hypomethylated "blocks" in CD4+ cells in WGBS, we observed a significant decrease in methylation in CD4+ samples from donors older than 80 as compared to CD4+ samples from donors 18-34 years of age (two sample t-test, $p < 2.2 \times 10^{-16}$) (Figure 2.6A). When we considered the small DMRs identified by WGBS, we were limited by the low coverage of the array within these regions, which are primarily far from gene promoters, so we were able to directly examine only a subset of DMRs. Among the 71 identified WGBS age-DMRs which overlap 3 or more probes on the 450k array (20 CD4+ age-DMRS, 47 nCD4+ age-DMRs, 4 CD14+ age-DMRs), 66 showed the same direction of differential methylation comparing samples from the youngest (18-34 years) and oldest (80+ donors) on the 450k array, 56 with a p-value <0.1 (Fig 2.6B).

We applied the "methylation age" model developed by Horvath[89] to determine how the model applies to sorted cell populations. When we compared the calculated methylation age to the chronological age, we observed a very strong correlation between methylation age and chronological age for all cell types (R^2's from 0.90 to 0.94). Notably, for all individuals, the methylation age for CD4+ cells is lower than the methylation age

of CD14+ cells and PBMCs, indicating cell type specificity in relationship between age and methylation even in this limited set of 353 cytosines (Figure 2.7A).

We next identified the probes most significantly differentially methylated with age. We fit a linear model for methylation versus age adjusting for sex and race, separately for each cell type and considered a probe an age-DMP for a given cell type if the relationship with age was significant at a level of $<5 \times 10^{-7}$. Consist with our observation in the WGBS data, most of the identified age-DMPs were cell type specific, with the largest number found in CD4+ cells (Figure 2.7B-E). Only one of the DMPs identified as common between all cell types is part of the Horvath methylation-aging model. However many of these common age-DMPs, including those near ELOVL2, KLF14 and FHL2 have been identified as methylation aging markers in other studies[155, 156]. Notably, many of the age-DMPs identified as cell type specific have similar nearest genes. For example, type specific age-DMPs in the gene body of PRDM16 are identified in each cell type, indicating that common genes may be regulated even if the highest magnitude of change occurs in different locations for each cell type.

Discussion

In this work, we demonstrate cell type specific patterns of methylation change with age in two types of peripheral blood cells, CD14+ monocytes and CD4+ T lymphocytes. We observe both genome scale change in the form of hypomethylated blocks and smaller regions of change, DMR. We demonstrate that the identified small DMRs overlap strongly with cell type specific enhancer regions and at least one set of DMRs is enriched for a motif at which a cell type regulatory factor binds. While recent work has focused on identifying aging related methylation changes that are common

between tissues or identifiable in mixed tissues like peripheral blood, our data demonstrates that some methylation change with age is cell type specific and should be studied in pure cell types in order to make mechanistic inferences.

We observe widespread hypomethylation with age accounted for by large hypomethylated blocks in CD4+ and naïve CD4+ lymphocytes, but not in monocytes. This is highly intriguing as it indicates that methylation in the large, CpG-sparse regions affected by blocks is preserved differently in each cell type. There are many factors which could contribute to this difference: CD4+ lymphocytes are part of the adaptive immune system, which develops and changes with aging, whilc CD14+ monocytes are part of the innate immune system which is less variable with age. CD4+ lymphocytes are present within peripheral blood for months to years [157], while CD14+ monocytes turnover much more frequently, with a lifespan of days[158], it is possible that loss of block methylation could accumulate as CD4+ cells sit in a quiescent state for long periods, while CD14+ cells do not stay long enough to accumulate this pattern. Additionally, CD14+ cells are not fully differentiated, they become macrophages or dendritic cells upon reaching their target tissue. It is possible that there is greater protection of global methylation in these cells than in the terminally differentiated CD4+ lymphocytes.

The pattern of small DMRs observed also strongly suggested differential methylation with age may be intimately linked to cell identity. Using a very stringent threshold, we observed a large number of age-DMRs in CD4+, naïve CD4+ and CD14+ cells. A small number of these age-DMRs are common between both cell types, indicating that there are likely common mechanisms of differential methylation present.

The fact that at least some of these regions are nearest to genes with known roles in aging, such as MIR30D and PIK3C2A, indicates that the common regions of change may play an important role in cellular aging and further work should be done to clarify the locations and mechanism underlying such change. However, the majority of regions identified were distinct between lymphocytes and monocytes, and the patterns of change were also distinct. Lymphocytes have more hypermethylated age-DMRs and monocytes have more hypo-methylated age DMRs. For lymphocytes, methylation aging closely resembles the changes observed in cancer sample: widespread hypomethylated blocks with local hypermethylation [145].

The strong enrichment for age-DMRs from both cell types in enhancers further supports the idea that methylation change with age is closely linked to the cells epigenomic identity. Despite the many differences in age-DMRs between monocytes and lymphocytes, for both cells types we observed a similar relationship between methylation and enhancers: hypomethylated age-DMRs overlap cell type specific enhancers, while enhancers from the opposite cell type are enriched in hypermethylated DMRs. Methylation patterns in enhancers have been linked to transcription factor binding [159], so it is possible that the observed age-DMRs are indicative of altered binding at cell type specific regulatory regions with age. It is striking that the methylation patterns suggest a more active binding environment at cell type specific enhancers with age. This could be indicative of increased canalization of cell types with age.

The observed enrichment of BATF::JUN consensus binding sequence within hypomethylated age-DMRs in naïve CD4+ cells illustrates this idea and demonstrates how identification of differentially methylated regions can identify important sites of

regulation. Binding at this motif in naïve CD4+ cells facilitates differentiation to Th17 cells. The balance and function of these cells has been observed to change with aging and may play a role in age related inflammatory phenotypes[154, 160]. The observation of hypomethylated DMRs in these binding sites provides insight into the stage and mechanism by which this pathway may be perturbed in aging.

Experimental Procedures

Human Subjects: 43 healthy donors took part in this study. All participants provided written informed consent, and this study was approved by the Johns Hopkins Medicine Institutional Review Board (IRB# NA_00052046). Screening criteria for donors less than 50 years included diagnosis of diabetes, thyroid disease, high blood pressure, depression or anxiety disorder, pregnancy, use of prescription medication other than birth control, use of recreational drugs and daily drinking. Screening criteria for donors over 50 years included rheumatoid arthritis, gout, COPD, asthma, heart attack within 6 months, Parkinson's disease, hip or knee surgery within 6 months, paralysis by stroke, active cancer or cancer diagnosis within 1 year, and use of steroid medications.

Cell sorting and DNA extraction: Peripheral blood mononuclear cells were isolated using Ficoll (Sigma-Aldrich). CD4+ T cells and CD14+ Monocytes were isolated through positive selection using MACS magnetic bead technology (Miltenyi Biotec). Naïve CD4 T+ Cells were isolated through selection for CD4 and CD45RA. Purity, as measured by post-separation flow cytometry was 97% for CD4+ cells, 95% for CD14+ cells and 85% for naïve CD4+ cells (Figure 2.1). Purified cells were flash frozen in liquid nitrogen and stored at -80°C.

Whole genome bisulfite sequencing: DNA was extracted from frozen cell pellets using the MasterPure DNA Purification kit (Epicentre) according the manufacturer's protocol. 1% unmethylated Lambda DNA (Promega, cat # D1521) was spiked in to genomic DNA to monitor the bisulfite conversion efficiency. Genomic DNA was fragmented to a target peak of 300-400 bp using the Covaris S2 Focused-ultrasonicator in a 50 µl volume according to the manufacturer's instructions.

The fragmented DNA was converted to end-repaired, adenylated DNA using the NEBNext Ultra End Repair/dA-Tailing Module (New England BioLabs, cat # 7442L). Methylated adaptors (NEBNext Multiplex Oligos for Illumina; New England BioLabs, cat # E7535L) were ligated to the product from the preceding step using the NEBNext Ultra Ligation Module (New England BioLabs, cat # 7445L). The resulting product was size-selected as described in the manufacturer's protocol by employing modified AMPure XP bead ratios of 0.4X and 0.2X in order to select for an insert size of 300-400 bp.

After size-selection the samples were bisulfite converted and purified using the EZ DNA Methylation- Gold Kit (Zymo Research, cat # D5005). Bisulfite converted libraries were PCR amplified and indexed using primers from the NEBNext Multiplex Oligos for Illumina module (New England BioLabs, cat # E7535L) and the Kapa HiFi Uracil+ PCR system (Kapa Biosystems, cat # KK2801). PCR enrichment was performed with the following cycling parameters: 98°C for 45 sec followed by 10 cycles at 98°C for 15 sec, 65°C for 30 sec, 72°C for 30 sec and a final extension at 72°C for 1 min. The PCR enriched product was cleaned up using 1X AMPure XP beads (Beckman Coulter, cat # A63881).

The resulting libraries were sequenced at a 2x100 bp read length on the Illumina HiSeq2000 platform using v3 chemistry according to the manufacturer's protocol.

Sequencing data analysis: All analyses were performed using R 3.0.1. To process sequencing data, we ran the BSmooth [45] bisulfite alignment pipeline on the 100-by-100 bp HiSeq 2000 paired end sequencing reads obtained for each sample, using Bowtie2 version 2.2.2 [142] and the hg19 build on the human genome as well as the genome for lambda phage. Table 2.2 summarizes the alignment results. After alignment, BSmooth was used to extract read-level measurements, summarized in Table 2.2. We filtered out measurements with mapping quality <20 or nucleotide base quality <10 and we removed measurements from the 5' most 10 nucleotides of both mates. BSmooth was used to sort read-level measurements by genomic coordinates and compile a summary table.

BSmooth was used to identify large hypomethylated blocks and small DMRs as described in detail previously [28, 31]. CpGs with coverage of 2 or greater in each cell type/age group were included in the analysis. We used the same cutoffs used in studies of cancer, specifically a t-statistic cutoff of ±2 for block finding and a t-statistic cutoff of ±4.6 for small DMR finding.

Chromatin peak information for naïve CD4 cells and monocytes was obtained from the Roadmap project [150], downloaded from egg2.wustl.edu/roadmap/wed_portal. Narrow peaks were used. Overlap with DMRs was calculated based on the odds ratio for each analyzed CpG occurring in each type of region.

Motif Finding: Identification of enriched motifs within DMR lists was performed using *Haystack*, as described in detail elsewhere [151]. Enrichment within hypo and

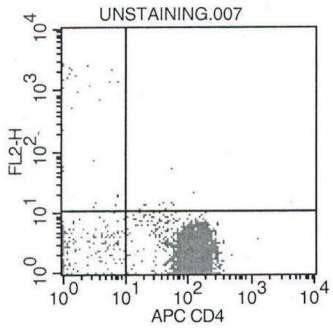
hyper methylated DMR lists for each cell type was compared to a background of all other DMRs.

Cytokine quantification: IL-6 was quantified in stored serum using Human IL-6 Quantikine ELISA kit (R&D systems) according to manufacturer's instructions. IL-17 was quantified in stored serum using Human IL-17A High Sensitivity ELISA kit (eBiosciences, San Diego, CA).

450k Array: DNA was quantified using Quant-iT Picogreen Reagent (Invitrogen) according to manufacturer's instructions. 500 ng of DNA was bisulfite converted using the EZ DNA methylation kit (Zymo Research) according to the manufacturer's specifications for the 450k array. Converted genomic DNA was eluted in 11 μl of elution buffer. DNA methylation level was measured using Illumina Infinium HD Methylation Assay (Illumina) according to the manufacturer's instructions.

450k Analysis: All analyses were performed using R 3.0.3. Raw intensity files were obtained and processed using the Minfi package [161] to obtain methylation ratios (Beta values). Samples were normalized using the Illumina preprocessing method implemented in Minfi. We applied multiple quality control measures to remove questionable arrays or probes. We examined 450k array control probes to assess many measures of assay efficiency and calculated median methylated and unmethylated measurements for each sample. We removed probes that had an annotated SNP (dbSNP137) at the single base extension or CpG site (17541 probes removed). Methylation age was calculated using 353 probe seen to be predictive of chronological age as detailed elsewhere [89]. Age associated differentially methylated probes were

identified using a linear model including race (white, black, other) and sex and adjusted for multiple testing using bonferroni correction.

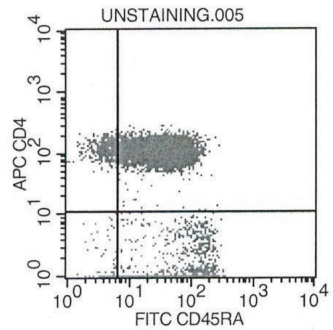


Quadrant Statistics

File: UNSTAINING.007
Sample ID:
Tube: CD4-donor 1
Acquisition Date: 25-Oct-13
Gated Events: 9460
X Parameter: FL4-H APC CD4 (Log)
Quad Location: 10, 11

Log Data Units: Linear Values
Patient ID: donor 3
Panel: HL-PBMC-purification
Gate: G1
Total Events: 10000
Y Parameter: FL2-H (Log)

Quad	Events	% Gated	% Total	X Mean	X Geo Mean	Y Mean	Y Geo Mean
UL	26	0.27	0.26	2.88	2.16	1112.16	401.90
UR	11	0.12	0.11	46.57	39.95	17.09	15.30
LL	221	2.34	2.21	2.55	1.82	3.08	2.63
LR	9202	97.27	92.02	125.30	119.11	2.13	1.86

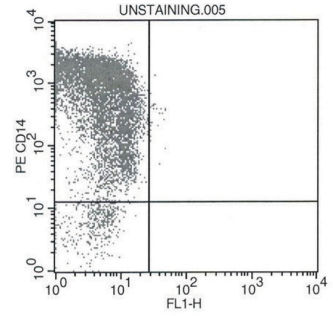


Quadrant Statistics

File: UNSTAINING.005
Sample ID:
Tube: CD4/CD45RA-donor1 N CD4
Acquisition Date: 25-Oct-13
Gated Events: 9673
X Parameter: FL1-H FITC CD45RA (Log)
Quad Location: 6, 11

Log Data Units: Linear Values
Patient ID: donor 3
Panel: HL-PBMC-purification
Gate: G1
Total Events: 10000
Y Parameter: FL4-H APC CD4 (Log)

Quad	Events	% Gated	% Total	X Mean	X Geo Mean	Y Mean	Y Geo Mean
UL	402	4.16	4.02	4.54	4.29	126.24	120.92
UR	8237	85.15	82.37	42.51	34.69	112.76	108.62
LL	59	0.61	0.59	4.32	4.06	2.59	1.88
LR	975	10.08	9.75	123.98	94.20	2.14	1.61



Quadrant Statistics

File: UNSTAINING.005
Patient ID:
Acquisition Date: 10-Apr-14
Gated Events: 8144
X Parameter: FL1-H (Log)
Quad Location: 27, 13

Log Data Units: Linear Values
Tube: PE-CD14
Gate: G1
Total Events: 10000
Y Parameter: FL2-H PE CD14 (Log)

Quad	Events	% Gated	% Total	X Mean	X Geo Mean	Y Mean	Y Geo Mean
UL	7713	94.71	77.13	6.28	4.48	1114.16	657.73
UR	26	0.32	0.26	35.19	34.53	348.92	276.13
LL	405	4.97	4.05	4.62	3.95	6.43	5.62
LR	0	0.00	0.00	***	***	***	***

Figure 2.1: Validation of magnetic bead purification.

Cell populations purified using magnetic beads were analyzed using flow cytometry to determine purity of CD4+ cells (top row), naïve CD4+ cells (middle row), CD14+ cells (bottom row).

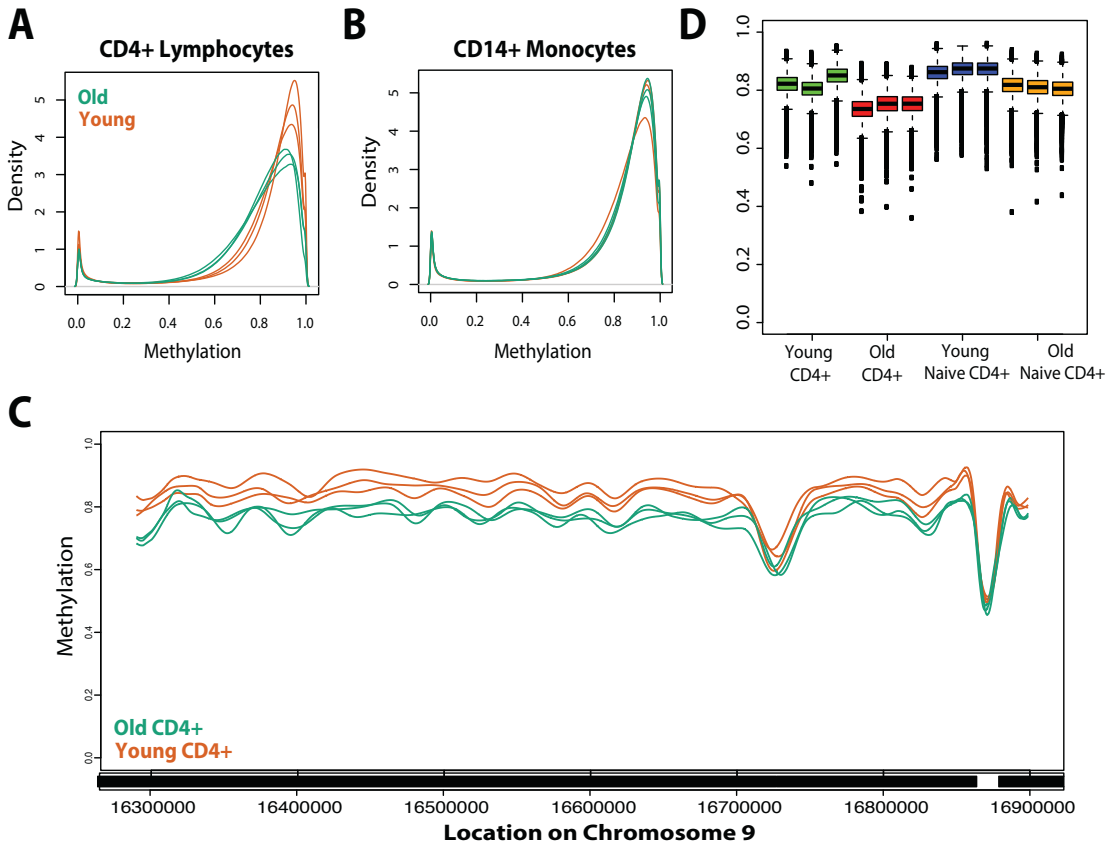


Figure 2.2: Widespread hypomethylated blocks in aged CD4+ cells.

A) Global hypomethylation with age in CD4+ lymphocytes. Distribution of high-frequency smoothed methylation values for CD4+ lymphocytes from CpGs with sufficient coverage from WGBS. Younger samples are shown in orange, older samples are shown in green. B) No global changes with age in CD14+ monocytes. Distribution of high-frequency smoothed methylation values for CD14+ monocytes from CpGs with sufficient coverage from WGBS. Younger samples are shown in orange, older samples are shown in green. C) Example of hypomethylated block identified by WGBS. Shown are smoothed methylation values within blocks identified comparing younger (orange) and older (green) CD4+ samples. D) Naïve CD4+ cells are hypomethylated with age in CD4+ age-blocks. Shown is the distribution of mean methylation within regions identified as age-blocks in CD4+ cells for CD4+ and naïve CD4+ samples from older and younger individuals.

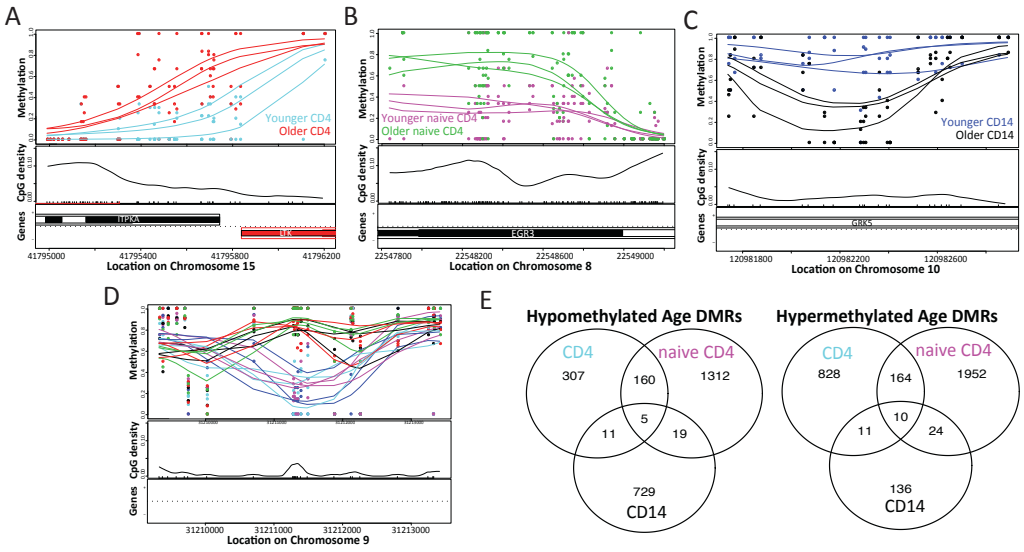


Figure 2.3: Cell type specific age-DMRs.

A) Example of hypermethylated CD4+ age-DMR identified by WGBS. Top panel: lines show smoothed methylation values within DMR identified comparing younger (blue) and older (red) CD4+ samples, points show raw methylation values. Middle Panel: CpG density across the identified region. Bottom panel: the location of genes within the identified region. Solid regions indicate exon locations. B) Example of hypermethylated naïve CD4+ age-DMR identified by WGBS. Top panel: lines show smoothed methylation values within DMR identified comparing younger (purple) and older (green) naive CD4+ samples, points show raw methylation values. Middle Panel: CpG density across the identified region. Bottom panel: the location of genes within the identified region. Solid regions indicate exon locations. C) Example of hypomethylated CD14+ age-DMR identified by WGBS. D) Example of common age DMR identified comparing older and younger samples in all three cell types. E) Hypomethylated age-DMRs are primarily cell type specific. Shown is a Venn diagram representing the number of regions identified as hypomethylated age-DMRs within each cell type. F) Hypermethylated age-DMRs are primarily cell type specific. Shown is a Venn diagram representing the number of regions identified as hypermethylated age-DMRs within each cell type.

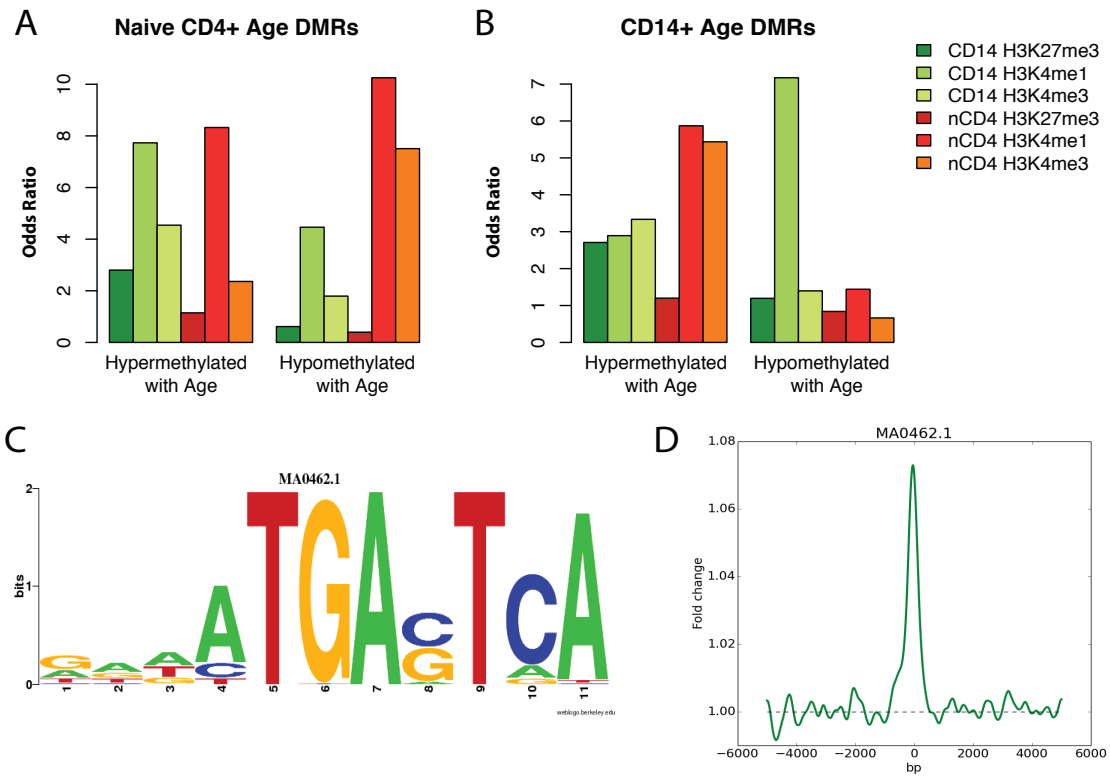


Figure 2.4: Age-DMRS are enriched for cell type specific enhancers and regulatory factor binding sites.

A) Naïve CD4+ age-DMRs are enriched for naïve CD4+ and CD14+ enhancer regions. Shown are odds-ratios for the overlap between hyper and hypomethylated naive CD4+ age DMRs and ChIP-seq peaks for each cell type. B) CD14+ age-DMRs are enriched for naïve CD4+ and CD14+ enhancer regions. Shown are odds-ratios for the overlap between hyper and hypomethylated naive CD4+ age DMRs and ChIP-seq peaks for each cell type. C) The BATF::JUN binding motif found to be enriched in naïve CD4+ age-DMRs. D) Center enrichment within naïve CD4+ age DMRs for the BATF::JUN binding motif.

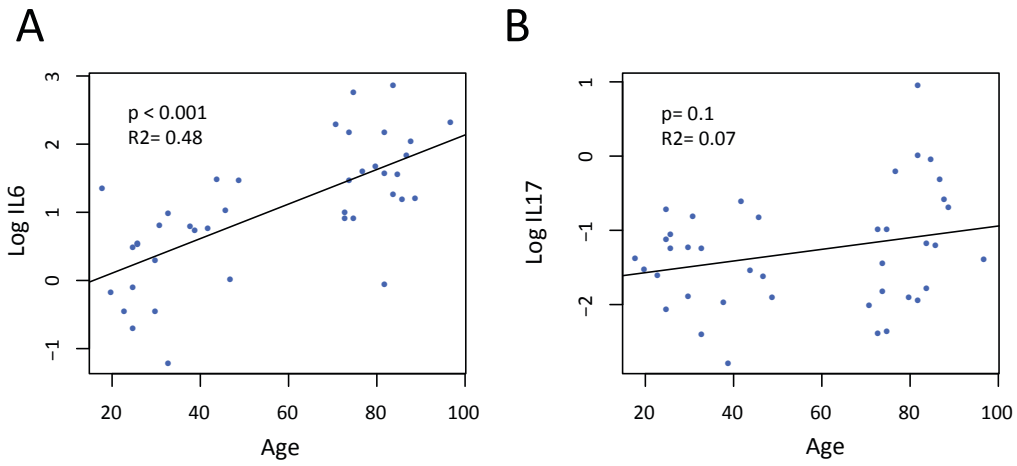


Figure 2.5: Cytokine levels in donor serum.

A) Log transformed IL-6 (pg/mL) measured by ELISA in stored serum from donors. B) Log transformed IL-17 (pg/mL) measured by ELISA in stored serum from donors.

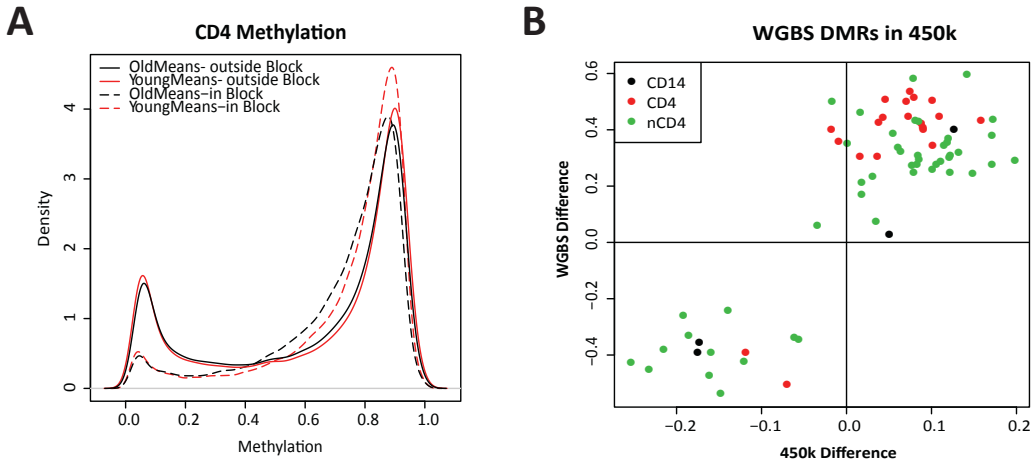


Figure 2.6: Validation of WGBS blocks and DMRs in 450k array data.

A) Shown is the distribution of mean CD4+ cell methylation from 450k array probes outside of regions identified as CD4+ age-blocks using WGBS (solid lines) or inside regions identified as CD4+ age-blocks using WGBS (dotted lines). Means all CD4+ samples from donors under 35 years shown in red, means from all CD4+ samples from donors over 80 years shown in black. B) Shown is the mean differential methylation measured using WGBS versus the mean differential methylation between samples from donors under 35 years and samples from donors over 80 years for regions identified as age-DMRs using WGBS which are covered by 3 or more probes in the 450k array. CD14+ age-DMRs are shown in black, CD4+ age-DMRs are shown in red, naïve CD4+ age-DMRs are shown in green.

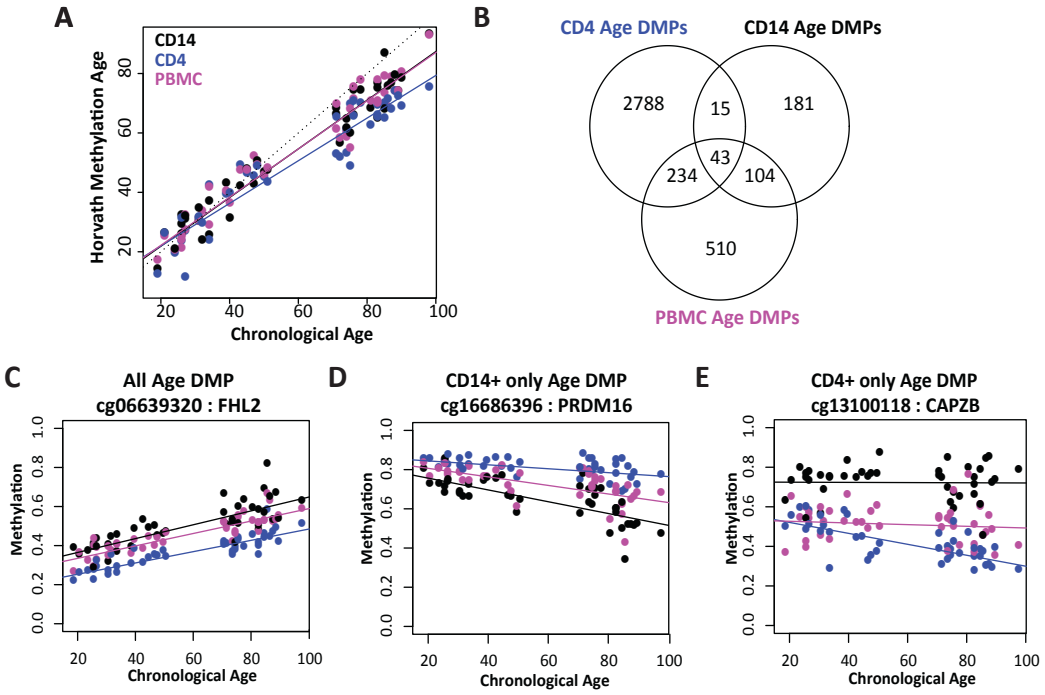


Figure 2.7: Cell type specific age-related methylation changes in 450k data.

A) Methylation age correlates with chronological age in all cell types. Shown is the "methylation age" calculated from 450k data using Horvath's algorithm versus the chronological age for each sample. CD14+ monocytes are shown in black, CD4+ lymphocytes are shown in blue, peripheral blood mononuclear cells are shown in pink. B) Age-DMPs are primarily cell type specific. Shown is a Venn diagram depicting the overlap of the most significant age-DMPs identified in each cell population. C) Example common age-DMP. Shown is methylation versus chronological age for a probe identified as a significant DMP for all cell types. D) Example CD14+ only age-DMP. Shown is methylation versus chronological age for a probe identified as a significant DMP for CD14+ cells only. E) Example CD4+ only age-DMP. Shown is methylation versus chronological age for a probe identified as a significant DMP CD4+ cells only.

Table 2.1: Donor Characteristics

ID	Age	Sex	Race	WGBS?
Y138	18	F	W	yes
Y140	20	M	B	no
Y171	23	F	O	no
Y134	25	F	B	no
Y147A	25	F	W	yes
Y172	25	F	W	yes
Y135	26	F	B	no
Y136	26	M	W	no
Y139	30	F	B	no
Y154	30	M	W	no
Y141	31	F	B	no
Y137	33	M	B	no
Y145	33	M	B	no
Y163	38	F	W	no
Y143	39	M	B	no
Y146	42	F	B	no
Y159	44	F	W	no
Y150	46	M	B	no
Y164	47	M	B	no
Y158	49	M	W	no
Y165	50	M	B	no
F943	70	F	W	no
F929	70	M	W	no
F955	71	F	W	no
F957	73	F	W	no
F971	73	M	W	no
F945	74	F	W	no
F965	74	M	W	no
F950	75	F	W	no
F948	75	F	W	no
F827A	77	M	O	no
F891	80	M	W	no
F879A	82	F	W	yes
F951	82	F	W	yes
F954	82	M	W	no
F920	84	F	W	no
F963	84	M	W	no
F927	85	F	B	no

ID	Age	Sex	Race	WGBS?
F600A	86	F	W	yes
F598A	87	F	W	no
F940	88	M	W	no
F944	89	M	W	no
F601A	97	F	W	no

Table 2.2:WGBS Summary

Sample	alignment rate	All CpGs	Covered CpGs	coverage	mean depth (X)	conversion
600A_14	0.833769908	28220561	25030852	0.89	5.46	0.9970
600A_4	0.834642824	28220561	24993780	0.89	5.11	0.9970
600A_n4	0.832945317	28220561	25131636	0.89	5.68	0.9972
819-14	0.835926723	28220561	25168034	0.89	5.75	0.9970
819-4	0.832601224	28220561	25260938	0.90	5.84	0.9969
819-n4	0.835227016	28220561	25223628	0.89	5.89	0.9971
951_4	0.835235647	28220561	25109823	0.89	5.98	0.9974
951_n4	0.835179283	28220561	25114414	0.89	5.97	0.9972
951-14	0.834172999	28220561	25192030	0.89	5.78	0.9971
Y138-14	0.838714195	28220561	25100392	0.89	5.75	0.9973
Y138-4	0.833807559	28220561	25176395	0.89	6.05	0.9973
Y138-n4	0.834665359	28220561	25113086	0.89	5.70	0.9973
Y147-14	0.83449785	28220561	25134611	0.89	5.69	0.9973
Y147-4	0.834247565	28220561	25125406	0.89	5.83	0.9973
Y147-n4	0.833491385	28220561	25150996	0.89	5.90	0.9974
Y172-14	0.835691499	28220561	25149409	0.89	5.83	0.9974
Y172-4	0.833406479	28220561	25210524	0.89	5.80	0.9973
Y172-n4	0.832528324	28220561	25234685	0.89	5.89	0.9970

Table 2.3: DMR finding summary

Cell Type	Hypo with age	Hyper with age	Mean Difference Hyper(min, max)	Mean Difference Hypo (min,max)
CD4	475	1012	-35% (-18%,-60%)	40% (16%,69%)
CD14	762	172	-37% (-20%,-61%)	37% (22%,61%)
nCD4	1495	2150	-30%(-3%,-77%)	35% (22%,63%)

Table 2.4: Common age-DMRs

Chr	Start	End	Nearest Gene	Distance from Nearest Gene	Direction of Change
chr1	105500662	105500779	AMY1B	1208200	hypermethylated
chr10	28231642	28231940	ARMC4	56036	hypermethylated
chr11	17195624	17195872	PIK3C2A	4271	hypermethylated
chr12	39363664	39364307	CPNE8	64245	hypermethylated
chr18	74310656	74310991	LOC284276	70045	hypermethylated
chr20	43312316	43312651	WISP2	31233	hypermethylated
chr6	122708685	122709030	HSF2	11665	hypermethylated
chr8	135826473	135826599	MIR30D	9286	hypermethylated
chr8	137048462	137048644	KHDRBS3	578747	hypermethylated
chr9	31211270	31211481	ACO1	1173119	hypermethylated
chr12	38284939	38285270	ALG10B	425286	hypomethylated
chr16	10794038	10794738	TEKT5	5237	hypomethylated
chr2	4981405	4981765	SOX11	851033	hypomethylated
chr2	101247565	101248116	PDCL3	68148	hypomethylated
chr21	35958435	35958587	RCAN1	28157	hypomethylated

Chapter 3: DNA methylation is stable during replication and cell cycle arrest

Introduction

DNA methylation is an epigenetic modification that has important functions in mammalian development. It consists of addition of a 5' methyl group to the cytosine base. This modification is most frequently found in the context of CpG dinucleotides, where it can be placed by three methyltransferase enzymes: DNMT1, the maintenance methyltransferase that re-establishes the methylation pattern following DNA replication, and DNMT3a and DNMT3b, which function in *de novo* methylation [6]. The presence of DNA methylation in gene promoters and enhancers decreases gene expression, likely through alterations of local DNA structure and prevention of transcription factor binding [143]. Changes in methylation are associated with aging, oncogenesis and other diseases [25, 89, 162-164].

Several molecular processes, including gene expression and chromatin structure, are known to change through the cell cycle [165-167]. These changes will result in unwanted variation between measures on bulk, unsynchronized cells. Computational approaches have been suggested to control for this variation [168, 169].

The extent to which DNA methylation changes throughout the cell cycle is currently unknown. Previous studies of methylation during the cell cycle have focused on the maintenance of methylation during DNA replication. The maintenance methyltransferase, DNMT1, localizes to newly synthesized DNA and is associated with the replication complex during S-phase and with other transcription factors during G0/G1 and G2/M phases [170]. A previous study in HeLa cells reported increased methylation during S phase using immunofluorescence and HPLC [171]. By contrast, a more recent study using flow cytometry to quantify 5-methylcytosine in two cancer cell lines found

no differences in the ratio of 5-methylcytosine staining to DNA content between G1 and S phase, but noted a lag in early G2/M before the maximum levels of 5mC were observed [172]. While these studies provide valuable insight into the methylation of newly synthesized DNA, both studies focus on changes during DNA synthesis, and neither provides region specific data.

We have previously reported the presence of large hypomethylated domains, encompassing up to two thirds of the genome, termed "hypomethylated blocks", in colon cancer samples and associated with EBV transformation of lymphocytes. These blocks occur in gene-poor regions and overlap strongly with heterochromatic and lamina-associated domains, indicating large-scale epigenetic structural changes [28, 31, 145]. These hypomethylated blocks also overlap strongly with regions of intermediate methylation in IMR90 cells [27]. A single cell can be methylated (100% methylation), hemimethylated (50% methylation), methylated in an allele specific manner (50% methylation) or unmethylated (0% methylation) at a given location, so the consistent observation of regions with intermediate methylation levels suggests that methylation in these areas are highly variable within cell populations. Intriguingly, large regions of hypomethylation are also observed in cells approaching replicative senescence [95]. In this context, hypomethylation was found in regions that replicate late in S phase, concurrent with decreased expression and decreased localization of DNMT1 to these domains in late passage cells late passage cells.

We noted that in both the cancer and EBV studies, hypomethylated blocks are seen in conditions where there is also an increase in the proportion of cells that are actively proliferating. Colon cancer is associated with increased cell proliferation outside

of the normal proliferative zone of colonic crypts as measured by Ki67 staining [173, 174] . EBV transformation directly promotes proliferation of previously resting cells by inducing expression of early G1 regulators [175]. Normal tissues are often distinguished by differences in the percent of proliferating cells, so any differences in methylation attributed to proliferation may be relevant to studies of tissue specific methylation as well [176]. Given these observations, we asked whether the observed hypomethylated blocks might be attributed to changes in the proportion of actively dividing cells within the populations studied. [170-172]

In this study, we sought to more fully elucidate the genome-scale changes in DNA methylation associated with cell proliferation. We used early passage primary dermal fibroblasts to avoid any artifacts from long-term cell culture and isolated quiescent and proliferating cells using flow cytometry. We further subdivided the proliferating cells into G1 and G2 phases to identify methylation changes as a result of DNA replication. We observed strikingly high degree of correlation between methylation across cell cycle phases within and between primary fibroblasts. We found no hypomethylated blocks or global changes in methylation associated with proliferation.

Results

No global changes in methylation associated with cell proliferation or replication

We sought to identify potential genome scale changes in DNA methylation associated with cell proliferation. Working with 3 sets of early passage (P4) primary human dermal fibroblasts (details in Table 3.1), we used fluorescence activated cell sorting to separate cells into three groups based on staining with anti-Ki67 and Propidium

Iodide: Quiescent (G0) cells were sorted based on negative Ki67 staining and 2N DNA and represented 46-71% of live cells; G1 cells were identified as Ki67 positive and 2N DNA and represented 5-21% of live cells; G2/M cells were identified as Ki67 positive and 4N DNA and represented 3-5% of live cells (Figure 3.1C). In addition, each set of primary cells was arrested by contact inhibition for 1 week to examine the influence of extended quiescence. After one week, 0.05-0.6% of all cells were in G2/M. Ki67 negative, 2N DNA cells were again isolated (Figure 3.1D).

In order to gain unbiased genome wide information about DNA methylation in these samples, we performed whole genome bisulfite sequencing (WGBS). We generated sequencing data to a depth of 5.3-8.6X and analyzed it using the BSmooth algorithm, which was designed for analyzing low-coverage WGBS data and has been demonstrated to accurately estimate methylation levels at single-base pair resolution by borrowing information from nearby CpGs [45]. After filtering reads with low quality measures, we obtained measurements for an average of 25,425,530 CpGs per sample (average of 90% coverage). Bisulfite conversion was assessed using spiked in lambda phage and ranged from 99.67-99.71% (Details in Methods, Table 3.2).

To identify potential large-scale changes in methylation associated with cell cycle stage, we analyzed data from all CpGs with two or more reads for at least two out of three of the samples for each cell cycle phase. The genome wide distribution of methylation was strikingly similar between samples within a cell cycle phase and between phases, with mean methylation ranging from 60-65% (shown in Figure 3.2A-C, summarized in Table 3.3). When we separately examined CpGs within the regions identified as hypomethylated blocks in colon cancer samples, we observed lower

methylation across all samples within colon cancer blocks as compared to outside, however no large differences were found between the cell cycle phases within or outside the blocks (Figure 3.2D).

When we compared mean methylation for each phase across all CpGs, we found a very strong correlation between methylation across all comparisons (Figure 3.3), indicating that large-scale differences observed in other settings are not found between cell cycle phases. Consistent with this observation, when we used BSmooth to identify potential methylation blocks within our dataset, none were identified using the criteria applied in previous work.

Discussion

In summary, we use low coverage whole genome bisulfite sequencing to demonstrate consistent levels of DNA methylation across phases of the cell cycle in human primary fibroblasts. We used flow sorting to isolate resting and actively dividing cells from low passage primary cell culture and also examined the same cells after one week of arrest. We observed a very strong correlation between methylation in each examined cell cycle phase within each donor's cells and between donors, with no evidence of genome wide change associated with either replication or quiescence.

This finding is informative for interpreting large-scale changes in methylation identified in other settings, including colon cancer, EBV immortalization and cellular senescence. We previously observed widespread hypomethylated blocks associated with colon cancer and EBV transformation [28, 31, 145]. Intriguingly, these hypomethylated blocks generally involve highly methylated regions shifting to intermediate levels of methylation, indicating that these regions are more variably methylated within the cell

population examined. Both of these conditions are associated with increased cell proliferation, and thus an increased percentage of Ki67+ and G2 cells within the population studied, which we hypothesized could explain such heterogeneity. However, our results indicate that these changes alone would not lead to the observed widespread hypomethylation. Further, while the results of a recent work by Cruikshanks et al. implicate proliferative history in the development of hypomethylated blocks [95], our results indicate that proliferation, as measured by Ki67 positivity or 4N DNA content, in low passage fibroblasts is not sufficient for this change. It is possible that our findings would be different in late passage cells, as Cruikshanks et al. indicate defects in DNMT1 localization only in late passage cells.

We examined cells that were arrested by contact inhibition for one week in order to determine if methylation changes would be introduced or accentuated by an extended exit from the cell cycle, similar to many somatic cell types *in vivo*. We did not observe any large-scale change in methylation associated with the extended cell cycle arrest. However, this length of contact inhibition is much shorter than the length of time many somatic cells spend without dividing, so it remains possible that global methylation changes observed *in vivo* may be linked to extended exit from the cell cycle.

Two previous studies examining global methylation levels across the cell cycle have focused on the dynamics of methylation maintenance during S phase, with an early work reporting hypermethylation during DNA replication and a contradictory recent work showing a linear relationship between DNA content and 5mC signal during replication [171, 172]. We did not examine S phase, as using partially replicated DNA would be particularly challenging to accurately measure methylation levels using low

coverage sequencing data. While Desjobert et al. report a lag in the time at which they detect maximum DNA content and maximum methylation signal in G2/M [172], our data shows no significant hypomethylation in G2/M cells, indicating that if such a lag exists, the duration and magnitude are not sufficient to be detected by our WGBS analysis.

Interpretation of our data is also limited by the use of a single cell type. We chose to use low passage primary fibroblasts to make our data relevant to the study of primary tissue samples. However these cells are less proliferative than ES or cancer cell lines and our data indicate that they have significant levels of intermediate methylation. It is possible that proliferation or cell cycle dependent changes may be present in other cell types.

Experimental Procedures

Tissue Culture: Primary dermal fibroblasts were maintained in Dulbecco's Modified Eagle Medium DMEM (Life Technologies) supplemented with 15% FBS (Gemini BioProducts) and 1X Pen/Strep (Gibco). Cells were incubated at 37°C with 5% CO_2. Donor information for each cell line is listed in Table S1.

Fluorescence-activated Cell Sorting: Passage 4 cells were detached using trypsin (Life Technologies). Trypsin was neutralized using trypsin neutralization buffer (ATCC), washed, fixed in cold 75% ethanol and stored at -20°C. Fixed cells were washed in PBS containing 1% FBS, 0.09% NaN_3. Cells were stained with FITC conjugated Mouse Anti-Human Ki67 (clone B56) and 25 ug/mL propidium iodide solution (Sigma). Fluorescence activated cell sorting analysis was performed using a Beckman Coulter MoFlo Cell Sorter.

Whole Genome Bisulfite Sequencing:1% unmethylated Lambda DNA (Promega, cat # D1521) was spiked in to genomic DNA to monitor the bisulfite conversion efficiency. 50-100 ng of genomic DNA was fragmented to a target peak of 300-400 bp using the Covaris S2 Focused-ultrasonicator in a 50 μl volume according to the manufacturer's instructions.

The fragmented DNA was converted to end-repaired, adenylated DNA using the NEBNext Ultra End Repair/dA-Tailing Module (New England BioLabs, cat # 7442L). Methylated adaptors (NEBNext Multiplex Oligos for Illumina; New England BioLabs, cat # E7535L) were ligated to the product from the preceding step using the NEBNext Ultra Ligation Module (New England BioLabs, cat # 7445L). The resulting product was size-selected as described in the manufacturer's protocol by employing modified AMPure XP bead ratios of 0.4X and 0.2X in order to select for an insert size of 300-400 bp.

After size-selection the samples were bisulfite converted and purified using the EZ DNA Methylation- Gold Kit (Zymo Research, cat # D5005). Bisulfite converted libraries were PCR amplified and indexed using primers from the NEBNext Multiplex Oligos for Illumina module (New England BioLabs, cat # E7535L) and the Kapa HiFi Uracil+ PCR system (Kapa Biosystems, cat # KK2801). PCR enrichment was performed with the following cycling parameters: 98°C for 45 sec followed by 10 cycles at 98°C for 15 sec, 65°C for 30 sec, 72°C for 30 sec and a final extension at 72°C for 1 min. The PCR enriched product was cleaned up using 1X AMPure XP beads (Beckman Coulter, cat # A63881).

The resulting libraries were sequenced at a 2x100 bp read length on the Illumina HiSeq2000 platform using v3 chemistry according to the manufacturer's protocol.

Data Analysis: All analyses were performed using R 3.0.1. To process sequencing data, we ran the BSmooth [24] bisulfite alignment pipeline on the 100-by-100 bp HiSeq 2000 paired end sequencing reads obtained for each sample, using Bowtie2 version 2.1.0 [43] and the hg19 build on the human genome as well as the genome for lambda phage. Supplementary Table S2 summarizes the alignment results. After alignment, BSmooth was used to extract read-level measurements, summarized in Supplementary Table S2. We filtered out measurements with mapping quality <20 or nucleotide base quality <10 and we removed measurements from the 5' most 10 nucleotides of both mates. BSmooth was used to sort read-level measurements by genomic coordinates and compile a summary table.

Next, BSmooth was used to identify large hypomethylated blocks as described in detail previously [17, 24, 25]. CpGs with coverage of 2 or greater in each sample group were included in the analysis. We used the same cutoffs used in studies of cancer, specifically a t-statistic cutoff of -2, 2.

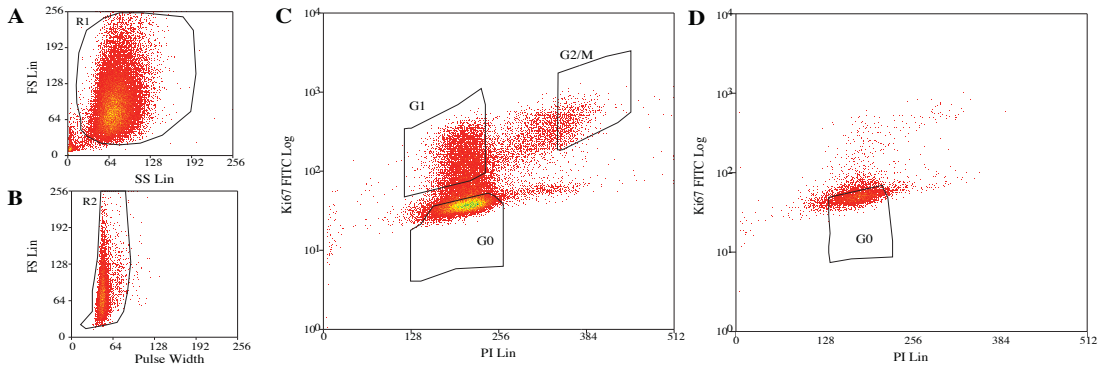


Figure 3.1: Sorting of fibroblasts based on Ki-67 expression and DNA content.

A) Live cells were selected using forward scatter and side scatter. B) Single cells were selected using pulse width. C) Cells from actively proliferating culture were sorted into "G0": 2N DNA and Ki67 negative, "G1": 2N DNA and Ki67 positive, "G2/M": 4N DNA and Ki67 positive. D) Cells from culture arrested by contact inhibition for one week were predominantly "G0".

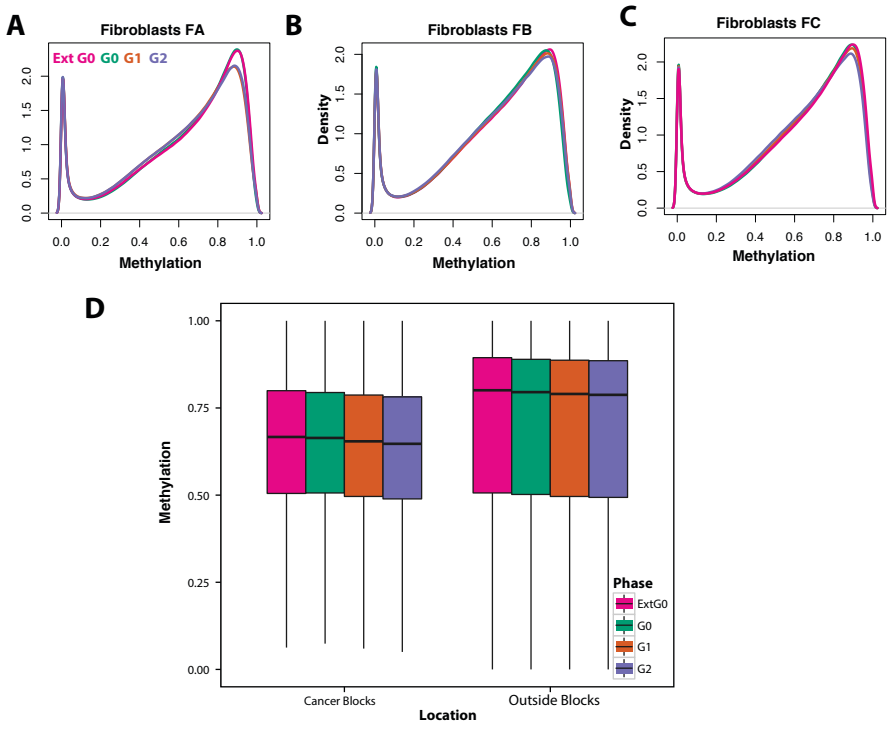


Figure 3.2: Global methylation is consistent after replication and cell cycle arrest.

(A-C) Distribution of high-frequency smoothed methylation values from CpGs with sufficient coverage from whole genome bisulfite sequencing (WGBS) for each donor profiled. (D) Distribution of mean high-frequency smoothed methylation values from CpGs within and outside the regions previously identified as hypomethylated blocks in colon cancer for each phase profiled. Outliers have been removed for plotting.

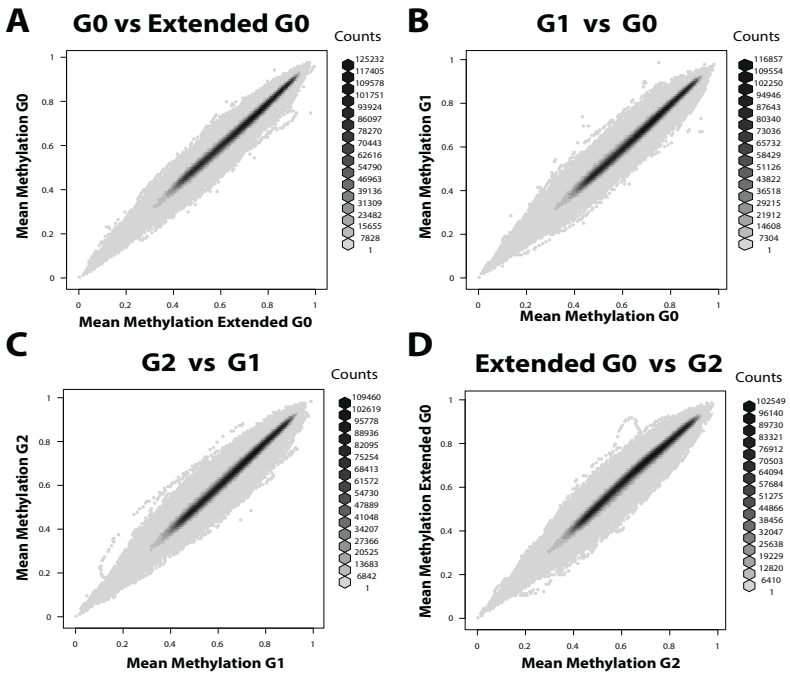


Figure 3.3: Global methylation is highly correlated between cell cycle phases.

(A-D) Shown is hexagonal binning of mean methylation per CpG for all samples in G0 versus all samples in extended G0 (A), all samples in G1 versus all samples in G0 (B), all samples in G2 versus all samples in G1 (C), all samples in extended G0 versus all samples in G2 (D).

Table 3.1: Primary cells analyzed

Cells	Supplier	Catalog #	Lot	Donor Age	Donor Sex	Donor Race
FA	Gibco	C-013-5C	1474560	36	Female	Caucasian
FB	Lonza	CC-2511	352805	40	Female	Caucasian
FC	ATCC	PCS-201-012	61447289	34	Female	Caucasian

Table 3.2: WGBS Summary

Sample	alignment rate	covered CpGs	coverage	mean depth (X)	conversion
FA_Ext_G0	0.8166	25514228	0.9041	7.2102	0.9970
FA_G0	0.8171	25647479	0.9088	8.6014	0.9970
FA_G1	0.8242	24826013	0.8797	5.2842	0.9971
FA_G2	0.8137	25322198	0.8973	6.6033	0.9971
FB_Ext_G0	0.8131	25350618	0.8983	6.2853	0.9971
FB_G0	0.8086	25393776	0.8998	6.9167	0.9970
FB_G1	0.8176	25312601	0.8970	5.6864	0.9968
FB_G2	0.8034	25313403	0.8970	5.6074	0.9971
FC_Ext_G0	0.8120	25647389	0.9088	7.5433	0.9970
FC_G0	0.8148	25665135	0.9094	7.4768	0.9970
FC_G1	0.8169	25602198	0.9072	7.1322	0.9969
FC_G2	0.8018	25511322	0.9040	6.5058	0.9968

Table 3.3: Mean Methylation

Methylation Mean *(1st, 3rd Quartile)*	**FA**	**FB**	**FC**
Extended G0	0.65 (0.40,1.00)	0.64 (0.40,1.00)	0.65 (0.40,1.00)
G0	0.65 (0.42,1.00)	0.63 (0.38,1.00)	0.65 (0.42,1.00)
G1	0.60 (0.40,1.00)	0.64 (0.38,1.00)	0.65 (0.40,1.00)
G2	0.64 (0.38,1.00)	0.63 (0.38,1.00)	0.64 (0.40,1.00)

Appendix: Hypomethylated blocks in long term passage of primary fibroblasts

Overview

Cellular senescence is a protective response in which a cell permanently exits the cell cycle in response to cell stress. It was first characterized *in vitro* by Hayflick who noted that primary fibroblasts have a limited replicative lifespan before exiting the cell cycle [177]. The response can also be induced in culture through serial passage, expression of oncogenes, or induction of genotoxic stress (reviewed elsewhere [53]).

The induction of senescence *in vitro* is characterized by specific cellular changes. The senescence response in serially passaged cells is partially attributed to shortening of telomere sequences, which induces a DNA damage response [178]. In addition to telomere shortening, senescent cells are characterized by a change in pH allowing for positive beta galactosidase staining [54], and epigenomic changes including a redistribution heterochromatin marks and loss of nuclear lamina proteins [179, 180].

Recent work by multiple groups indicates that DNA methylation is altered with extended culture of primary cells. Bork et al. have identified specific CpGs differentially methylated with replicative and stress induced senescence in mesenchymal stromal cells [181]. Horvath reports an increase in "methylation age" with serial passage of cultured fibroblasts [89]. Work by Cruikshanks et al. identified widespread hypomethylated blocks, similar to those observed in colon cancer, associated with replicative senescence in IMR-90 cells [95]. We cultured primary neonatal dermal fibroblasts to the point of growth arrest to further examine methylation changes associated with extended culture and replicative senescence and to determine the relevance of this system as a *in vitro* model of methylation aging.

Serial passage of primary fibroblasts

In order to investigate the effects of serial passage on primary cells, primary neonatal dermal fibroblasts (Lonza) were cultured in DMEM (Gibco) with 15% FBS (Gemini BioProducts) until there was no population doubling after 2 weeks, which occurred at passage 33. To verify that the growth arrest was consistent with replicative senescence, telomere length from early passage (passage 7) and senescent (passage 33) cells was measured by qPCR, comparing the amount of telomere repeat sequence compared to a single copy locus, 36B4, as detailed by Cawthorn et al. [182]. A significant decrease in the telomeric: single copy ratio was observed in passage 33 cells (average ratio 0.48) compared to passage 7 cells (average ratio 1.38).

450k analysis of methylation

Genomic DNA was extracted from passage 7 and passage 33 cells. Three technical replicates of DNA from each passage was bisulfite converted and genome wide methylation was analyzed using the 450k array. Array data was processed using *Minfi* and normalized using the preprocess Illumina method [110]. Consistent with Cruikshanks et al.'s observations, application of *Minfi*'s blockfinder identified large hypomethylated blocks in passage 37 cells, encompassing up to 235 Mbp (examples in Figure A.1A – A.1B).

WGBS analysis of methylation

To better characterize the pattern and timing at which widespread hypomethylated blocks develop during serial passage, genomic DNA from cells at the earliest available stage of culture (passage 4), cells at passage 7, cells at an intermediate stage of culture

(passage 10), cells approaching replicative senescence (passage 31) and cells at the final passage (passage 33) was analyzed using WGBS. We generated sequencing data to a depth of 8.0-8.9X and analyzed it using the BSmooth algorithm [45]. After filtering reads with low quality measures, we obtained measurements for an average of 25,749,493 CpGs per sample (average of 91% coverage) (summarized in Table A.1). Considering CpGs with >2 coverage in all samples, global methylation was observed to decrease progressively with each stage of passage (Table A.1). Consistent with our observations from 450k and the work of Cruikshanks et al., the majority of hypomethylation was observed within regions identified as hypomethylated blocks in colon cancer (Figure A.1C-A.1D). Notably, global and block methylation is similar between passage 31 (when culture was still actively proliferating) and passage 33 (when the culture failed to double), suggesting the observed change may not be specific to cell cycle arrest.

Summary

We observe widespread genomic blocks of hypomethylation in primary fibroblasts cultured to replicative senescence, consistent with previously reported methylation changes with senescence. The similarity between the changes observed in this system and those observed in cancer and aging suggest a possible link between a cell's replicative history and the development of large-scale epigenetic change. This finding provides a useful system for examining both the mechanism through which widespread methylation changes occur and the functional consequences of these changes in future work.

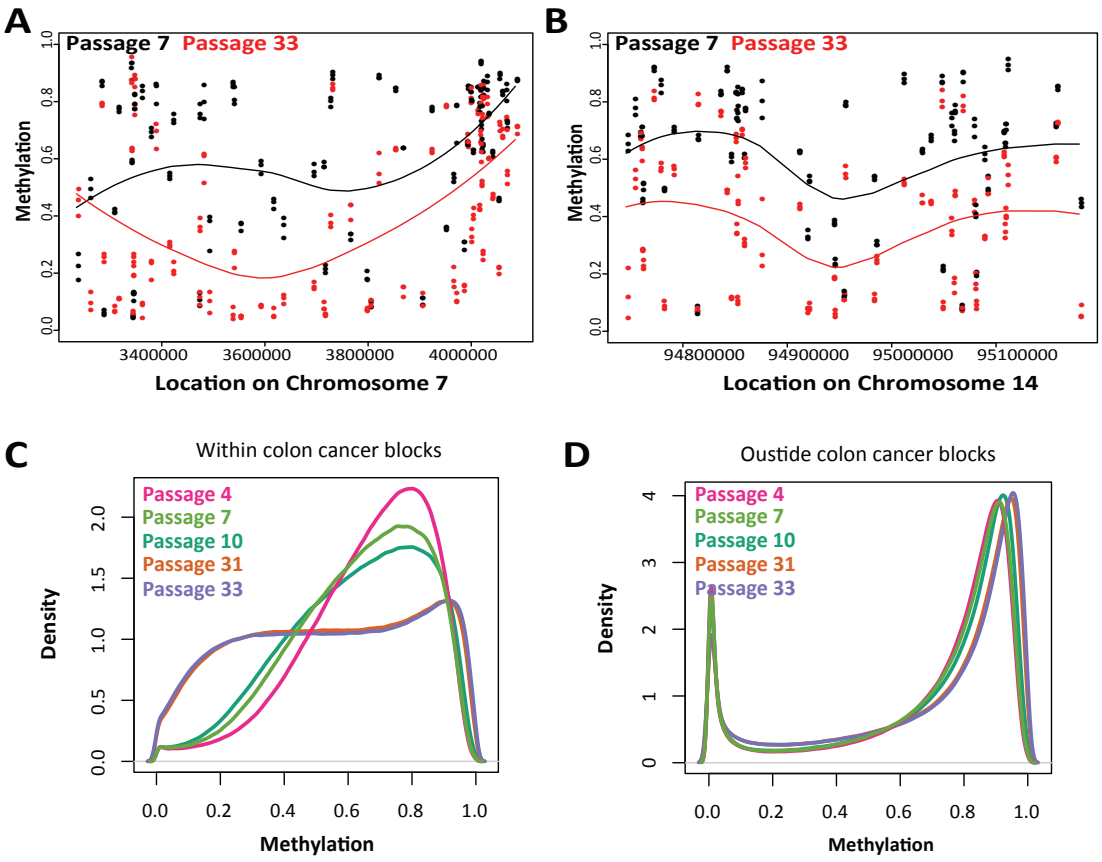


Figure A.1: Hypomethylated blocks in serially passaged fibroblasts.

A-B) Example of a regions identified as a block comparing passage 7 (black) and passage 37 (red) fibroblasts using 450k data. Shown are methylation beta values (methylated signal/total signal) for "collapsed" measurements of methylation from open sea probes in 450k data. These are methylation averages for each 1500 bp open sea region calculated as part of the Minfi's "block finder" algorithm. The points represent individual samples at each location, solid lines represent the smoothed average for each group. C) Distribution of high-frequency smoothed methylation values from CpGs within regions previously identified as hypomethylated blocks in colon cancer [28] with sufficient coverage from whole genome bisulfite sequencing (WGBS) for each sample analyzed. Methylation from passage 4 cells is depicted in pink, from passage 7 cells in light green, from passage 10 cells in dark green, form passage 31 cells in orange, from passage 33 cells in purple. D) Distribution of high-frequency smoothed methylation values from CpGs outside of regions previously identified as hypomethylated blocks in colon cancer [28] with sufficient coverage from whole genome bisulfite sequencing (WGBS) for each sample analyzed. Colored as in (C).

Table A.1: WGBS Summary

Sample	alignment rate	all CpGs	covered CpGs	coverage	mean depth (X)	mean methylation
Passage 4	0.8226	28220561	25826157	0.9152	8.9000	67.9%
Passage 7	0.8260	28220561	25786408	0.9137	8.4778	66.2%
Passage 10	0.8250	28220561	25720034	0.9114	7.9993	65.8%
Passage 31	0.8227	28220561	25766420	0.9130	8.4733	60.3%
Passage 33	0.8181	28220561	25648444	0.9089	7.9409	60.5%

References

1. Van Speybroeck L: **From epigenesis to epigenetics: the case of C. H. Waddington.** *Ann N Y Acad Sci* 2002, **981:**61-81.
2. Ziller MJ, Muller F, Liao J, Zhang Y, Gu H, Bock C, Boyle P, Epstein CB, Bernstein BE, Lengauer T, et al: **Genomic distribution and inter-sample variation of non-CpG methylation across human cell types.** *PLoS Genet* 2011, **7:**e1002389.
3. Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, et al: **Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain.** *Nat Neurosci* 2014, **17:**215-222.
4. Gardiner-Garden M, Frommer M: **CpG islands in vertebrate genomes.** *J Mol Biol* 1987, **196:**261-282.
5. Smith ZD, Meissner A: **DNA methylation: roles in mammalian development.** *Nat Rev Genet* 2013, **14:**204-220.
6. Okano M, Xie S, Li E: **Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases.** *Nat Genet* 1998, **19:**219-220.
7. Hervouet E, Vallette FM, Cartron PF: **Dnmt1/Transcription factor interactions: an alternative mechanism of DNA methylation inheritance.** *Genes Cancer* 2010, **1:**434-443.
8. Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA: **Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements.** *Mol Cell Biol* 2002, **22:**480-491.
9. Gopalakrishnan S, Sullivan BA, Trazzi S, Della Valle G, Robertson KD: **DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions.** *Hum Mol Genet* 2009, **18:**3178-3193.
10. Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D: **Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation.** *Nature* 2015.

11. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y: **Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.** *Nature* 2010, **466:**1129-1133.

12. Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, Blau HM: **Reprogramming towards pluripotency requires AID-dependent DNA demethylation.** *Nature* 2010, **463:**1042-1047.

13. Kagiwada S, Kurimoto K, Hirota T, Yamaji M, Saitou M: **Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice.** *EMBO J* 2013, **32:**340-353.

14. Yamaguchi S, Hong K, Liu R, Shen L, Inoue A, Diep D, Zhang K, Zhang Y: **Tet1 controls meiosis by regulating meiotic gene expression.** *Nature* 2012, **492:**443-447.

15. Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA: **Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine.** *Science* 2013, **339:**448-452.

16. Watt F, Molloy PL: **Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter.** *Genes Dev* 1988, **2:**1136-1143.

17. Gaston K, Fried M: **CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes.** *Nucleic Acids Res* 1995, **23:**901-909.

18. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A: **Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex.** *Nature* 1998, **393:**386-389.

19. Nan X, Campoy FJ, Bird A: **MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin.** *Cell* 1997, **88:**471-481.

20. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, Wei CL: **Dynamic changes in the human methylome during differentiation.** *Genome Res* 2010, **20:**320-331.

21. Gonzalo S, Jaco I, Fraga MF, Chen T, Li E, Esteller M, Blasco MA: **DNA methyltransferases control telomere length and telomere recombination in mammalian cells.** *Nat Cell Biol* 2006, **8:**416-424.

22. Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E: **Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene.** *Nature* 1999, **402:**187-191.

23. Muotri AR, Marchetto MC, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH: **L1 retrotransposition in neurons is modulated by MeCP2.** *Nature* 2010, **468:**443-446.

24. Jahner D, Stuhlmann H, Stewart CL, Harbers K, Lohler J, Simon I, Jaenisch R: **De novo methylation and expression of retroviral genomes during mouse embryogenesis.** *Nature* 1982, **298:**623-628.

25. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, et al: **The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.** *Nat Genet* 2009, **41:**178-186.

26. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, et al: **Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.** *Nat Genet* 2009, **41:**1350-1353.

27. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 2009, **462:**315-322.

28. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, et al: **Increased methylation variation in epigenetic domains across cancer types.** *Nat Genet* 2011, **43:**768-775.

29. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B: **Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.** *Nature* 2008, **453:**948-951.

30. Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP: **Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells.** *Nat Genet* 2009, **41:**246-250.

31. Hansen KD, Sabunciyan S, Langmead B, Nagy N, Curley R, Klein G, Klein E, Salamon D, Feinberg AP: **Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization.** *Genome Res* 2014, **24:**177-184.

32. Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, et al: **Comprehensive methylome map of lineage commitment from haematopoietic progenitors.** *Nature* 2010, **467:**338-342.

33. Trowbridge JJ, Snow JW, Kim J, Orkin SH: **DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells.** *Cell Stem Cell* 2009, **5:**442-449.

34. Sen GL, Reuter JA, Webster DE, Zhu L, Khavari PA: **DNMT1 maintains progenitor function in self-renewing somatic tissue.** *Nature* 2010, **463:**563-567.

35. Sheaffer KL, Kim R, Aoki R, Elliott EN, Schug J, Burger L, Schubeler D, Kaestner KH: **DNA methylation is required for the control of stem cell differentiation in the small intestine.** *Genes Dev* 2014, **28:**652-664.

36. Challen GA, Sun D, Mayle A, Jeong M, Luo M, Rodriguez B, Mallaney C, Celik H, Yang L, Xia Z, et al: **Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells.** *Cell Stem Cell* 2014, **15:**350-364.

37. Nishikawa K, Iwamoto Y, Kobayashi Y, Katsuoka F, Kawaguchi S, Tsujita T, Nakamura T, Kato S, Yamamoto M, Takayanagi H, Ishii M: **DNA methyltransferase 3a regulates osteoclast differentiation by coupling to an S-adenosylmethionine-producing metabolic pathway.** *Nat Med* 2015, **21:**281-287.

38. Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L, Schubeler D: **Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions.** *PLoS Genet* 2013, **9:**e1003994.

39. Epsztejn-Litman S, Feldman N, Abu-Remaileh M, Shufaro Y, Gerson A, Ueda J, Deplus R, Fuks F, Shinkai Y, Cedar H, Bergman Y: **De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes.** *Nat Struct Mol Biol* 2008, **15:**1176-1183.

40. Myant K, Stancheva I: **LSH cooperates with DNA methyltransferases to repress transcription.** *Mol Cell Biol* 2008, **28:**215-226.

41. Irizarry RA, Ladd-Acosta C, Carvalho B, Wu H, Brandenburg SA, Jeddeloh JA, Wen B, Feinberg AP: **Comprehensive high-throughput arrays for relative methylation (CHARM).** *Genome Res* 2008, **18:**780-790.

42. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL: **Genome-wide DNA methylation profiling using Infinium(R) assay.** *Epigenomics* 2009, **1:**177-200.

43. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M: **Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome.** *Epigenetics* 2011, **6:**692-702.

44. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL: **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.** *Proc Natl Acad Sci U S A* 1992, **89:**1827-1831.

45. Hansen KD, Langmead B, Irizarry RA: **BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions.** *Genome Biol* 2012, **13:**R83.

46. Niccoli T, Partridge L: **Ageing as a risk factor for disease.** *Curr Biol* 2012, **22:**R741-752.

47. Sousa-Victor P, Gutarra S, Garcia-Prat L, Rodriguez-Ubreva J, Ortet L, Ruiz-Bonilla V, Jardi M, Ballestar E, Gonzalez S, Serrano AL, et al: **Geriatric muscle stem cells switch reversible quiescence into senescence.** *Nature* 2014, **506:**316-321.

48. Keyes BE, Segal JP, Heller E, Lien WH, Chang CY, Guo X, Oristian DS, Zheng D, Fuchs E: **Nfatc1 orchestrates aging in hair follicle stem cells.** *Proc Natl Acad Sci U S A* 2013, **110:**E4950-4959.

49. Quarto R, Thomas D, Liang CT: **Bone progenitor cell deficits and the age-associated decline in bone repair capacity.** *Calcif Tissue Int* 1995, **56:**123-129.

50. Park SK, Prolla TA: **Gene expression profiling studies of aging in cardiac and skeletal muscles.** *Cardiovasc Res* 2005, **66:**205-212.

51. Varani J, Dame MK, Rittie L, Fligiel SE, Kang S, Fisher GJ, Voorhees JJ: **Decreased collagen production in chronologically aged skin: roles of age-dependent alteration in fibroblast function and defective mechanical stimulation.** *Am J Pathol* 2006, **168:**1861-1868.

52. Fagiolo U, Cossarizza A, Scala E, Fanales-Belasio E, Ortolani C, Cozzi E, Monti D, Franceschi C, Paganelli R: **Increased cytokine production in mononuclear cells of healthy elderly people.** *Eur J Immunol* 1993, **23:**2375-2378.

53. Campisi J, d'Adda di Fagagna F: **Cellular senescence: when bad things happen to good cells.** *Nat Rev Mol Cell Biol* 2007, **8:**729-740.

54. Dimri GP, Lee X, Basile G, Acosta M, Scott G, Roskelley C, Medrano EE, Linskens M, Rubelj I, Pereira-Smith O, et al.: **A biomarker that identifies senescent human cells in culture and in aging skin in vivo.** *Proc Natl Acad Sci U S A* 1995, **92:**9363-9367.

55. Melk A, Schmidt BM, Takeuchi O, Sawitzki B, Rayner DC, Halloran PF: **Expression of p16INK4a and other cell cycle regulator and senescence associated genes in aging human kidney.** *Kidney Int* 2004, **65:**510-520.

56. Chimenti C, Kajstura J, Torella D, Urbanek K, Heleniak H, Colussi C, Di Meglio F, Nadal-Ginard B, Frustaci A, Leri A, et al: **Senescence and death of primitive cells and myocytes lead to premature cardiac aging and heart failure.** *Circ Res* 2003, **93:**604-613.

57. Krtolica A, Parrinello S, Lockett S, Desprez PY, Campisi J: **Senescent fibroblasts promote epithelial cell growth and tumorigenesis: a link between cancer and aging.** *Proc Natl Acad Sci U S A* 2001, **98:**12072-12077.

58. Rodier F, Coppe JP, Patil CK, Hoeijmakers WA, Munoz DP, Raza SR, Freund A, Campeau E, Davalos AR, Campisi J: **Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion.** *Nat Cell Biol* 2009, **11:**973-979.

59. Baker DJ, Wijshake T, Tchkonia T, LeBrasseur NK, Childs BG, van de Sluis B, Kirkland JL, van Deursen JM: **Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders.** *Nature* 2011, **479:**232-236.

60. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G: **The hallmarks of aging.** *Cell* 2013, **153:**1194-1217.

61. Jin C, Li J, Green CD, Yu X, Tang X, Han D, Xian B, Wang D, Huang X, Cao X, et al: **Histone demethylase UTX-1 regulates C. elegans life span by targeting the insulin/IGF-1 signaling pathway.** *Cell Metab* 2011, **14:**161-172.

62. Larson K, Yan SJ, Tsurumi A, Liu J, Zhou J, Gaur K, Guo D, Eickbush TH, Li WX: **Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis.** *PLoS Genet* 2012, **8:**e1002473.

63. Greer EL, Maures TJ, Hauswirth AG, Green EM, Leeman DS, Maro GS, Han S, Banko MR, Gozani O, Brunet A: **Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in C. elegans.** *Nature* 2010, **466:**383-387.

64. Siebold AP, Banerjee R, Tie F, Kiss DL, Moskowitz J, Harte PJ: **Polycomb Repressive Complex 2 and Trithorax modulate Drosophila longevity and stress resistance.** *Proc Natl Acad Sci U S A* 2010, **107:**169-174.

65. Maures TJ, Greer EL, Hauswirth AG, Brunet A: **The H3K27 demethylase UTX-1 regulates C. elegans lifespan in a germline-independent, insulin-dependent manner.** *Aging Cell* 2011, **10:**980-990.

66. Hoath SB, Leahy DG: **The organization of human epidermis: functional epidermal units and phi proportionality.** *J Invest Dermatol* 2003, **121:**1440-1446.

67. Kaidbey KH, Agin PP, Sayre RM, Kligman AM: **Photoprotection by melanin--a comparison of black and Caucasian skin.** *J Am Acad Dermatol* 1979, **1:**249-260.

68. Bustamante J, Bredeston L, Malanga G, Mordoh J: **Role of melanin as a scavenger of active oxygen species.** *Pigment Cell Res* 1993, **6:**348-353.

69. Zouboulis CC, Makrantonaki E: **Clinical aspects and molecular diagnostics of skin aging.** *Clin Dermatol* 2011, **29:**3-14.

70. El-Domyati M, Attia S, Saleh F, Brown D, Birk DE, Gasparro F, Ahmad H, Uitto J: **Intrinsic aging vs. photoaging: a comparative histopathological, immunohistochemical, and ultrastructural study of skin.** *Exp Dermatol* 2002, **11:**398-405.

71. Fisher GJ, Datta SC, Talwar HS, Wang ZQ, Varani J, Kang S, Voorhees JJ: **Molecular basis of sun-induced premature skin ageing and retinoid antagonism.** *Nature* 1996, **379:**335-339.

72. Ghadially R, Brown BE, Sequeira-Martin SM, Feingold KR, Elias PM: **The aged epidermal permeability barrier. Structural, functional, and lipid biochemical abnormalities in humans and a senescent murine model.** *J Clin Invest* 1995, **95:**2281-2290.

73. Sextius P, Marionnet C, Tacheau C, Bon FX, Bastien P, Mauviel A, Bernard BA, Bernerd F, Dubertret L: **Analysis of gene expression dynamics revealed delayed and abnormal epidermal repair process in aged compared to young skin.** *Arch Dermatol Res* 2015.

74. Ashcroft GS, Mills SJ, Ashworth JJ: **Ageing and wound healing.** *Biogerontology* 2002, **3:**337-345.

75. Gloster HM, Jr., Brodland DG: **The epidemiology of skin cancer.** *Dermatol Surg* 1996, **22:**217-226.

76. Linton PJ, Dorshkind K: **Age-related changes in lymphocyte development and function.** *Nat Immunol* 2004, **5:**133-139.

77. Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ, Weissman IL: **Cell intrinsic alterations underlie hematopoietic stem cell aging.** *Proc Natl Acad Sci U S A* 2005, **102:**9194-9199.

78. Pang WW, Price EA, Sahoo D, Beerman I, Maloney WJ, Rossi DJ, Schrier SL, Weissman IL: **Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age.** *Proc Natl Acad Sci U S A* 2011, **108:**20012-20017.

79. Lazuardi L, Jenewein B, Wolf AM, Pfister G, Tzankov A, Grubeck-Loebenstein B: **Age-related loss of naive T cells and dysregulation of T-cell/B-cell interactions in human lymph nodes.** *Immunology* 2005, **114:**37-43.

80. Walston JD, Matteini AM, Nievergelt C, Lange LA, Fallin DM, Barzilai N, Ziv E, Pawlikowska L, Kwok P, Cummings SR, et al: **Inflammation and stress-related candidate genes, plasma interleukin-6 levels, and longevity in older adults.** *Exp Gerontol* 2009, **44:**350-355.

81. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P, Baccarelli A: **Decline in genomic DNA methylation through aging in a cohort of elderly subjects.** *Mech Ageing Dev* 2009, **130:**234-239.

82. Golbus J, Palella TD, Richardson BC: **Quantitative changes in T cell DNA methylation occur during differentiation and ageing.** *Eur J Immunol* 1990, **20:**1869-1872.

83. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstrom TJ, Harris TB, et al: **Intra-individual change over time in DNA methylation with familial clustering.** *JAMA* 2008, **299:**2877-2883.

84. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, et al: **Epigenetic differences arise during the lifetime of monozygotic twins.** *Proc Natl Acad Sci U S A* 2005, **102:**10604-10609.

85. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, et al: **Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains.** *Genome Res* 2010, **20:**434-439.

86. Zykovich A, Hubbard A, Flynn JM, Tarnopolsky M, Fraga MF, Kerksick C, Ogborn D, MacNeil L, Mooney SD, Melov S: **Genome-wide DNA methylation changes with age in disease-free human skeletal muscle.** *Aging Cell* 2014, **13:**360-366.

87. Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, Wininger M, Colantuoni C, Weinberger DR, Kleinman JE, Lipska BK: **DNA methylation signatures in development and aging of the human prefrontal cortex.** *Am J Hum Genet* 2012, **90:**260-272.

88. Fernandez AF, Bayon GF, Urdinguio RG, Torano EG, Garcia MG, Carella A, Petrus-Reurer S, Ferrero C, Martinez-Camblor P, Cubillo I, et al: **H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells.** *Genome Res* 2015, **25:**27-40.

89. Horvath S: **DNA methylation age of human tissues and cell types.** *Genome Biol* 2013, **14:**R115.

90. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, et al: **DNA methylation age of blood predicts all-cause mortality in later life.** *Genome Biol* 2015, **16:**25.

91. Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, Gibson J, Redmond P, Cox SR, Pattie A, et al: **The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936.** *Int J Epidemiol* 2015.

92. Saxena PK: **On the occurrence of the pretrematic branch of nervus glossopharyngeus in certain Indian teleostean fishes.** *Anat Anz* 1969, **124:**574.

93. Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D, Di Blasio AM, Giuliani C, Tung S, Vinters HV, Franceschi C: **Accelerated epigenetic aging in Down syndrome.** *Aging Cell* 2015.

94. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ, et al: **Distinct DNA methylomes of newborns and centenarians.** *Proc Natl Acad Sci U S A* 2012, **109:**10522-10527.

95. Cruickshanks HA, McBryan T, Nelson DM, Vanderkraats ND, Shah PP, van Tuyn J, Singh Rai T, Brock C, Donahue G, Dunican DS, et al: **Senescent cells harbour features of the cancer epigenome.** *Nat Cell Biol* 2013, **15:**1495-1506.

96. Gronniger E, Weber B, Heil O, Peters N, Stab F, Wenck H, Korn B, Winnefeld M, Lyko F: **Aging and chronic sun exposure cause distinct epigenetic changes in human skin.** *PLoS Genet* 2010, **6:**e1000971.

97. Raddatz G, Hagemann S, Aran D, Sohle J, Kulkarni PP, Kaderali L, Hellman A, Winnefeld M, Lyko F: **Aging is associated with highly defined epigenetic changes in the human epidermis.** *Epigenetics Chromatin* 2013, **6:**36.

98. Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, Burke G, Post W, Shea S, Jacobs DR, Jr., et al: **Age-related variations in the methylome associated with gene expression in human monocytes and T cells.** *Nat Commun* 2014, **5:**5366.

99. Beerman I, Bock C, Garrison BS, Smith ZD, Gu H, Meissner A, Rossi DJ: **Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging.** *Cell Stem Cell* 2013, **12:**413-425.

100. Nebel A, Kleindorp R, Caliebe A, Nothnagel M, Blanche H, Junge O, Wittig M, Ellinghaus D, Flachsbart F, Wichmann HE, et al: **A genome-wide association study confirms APOE as the**

major gene influencing survival in long-lived individuals. *Mech Ageing Dev* 2011, **132:**324-330.

101. Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, Kremer D, van der Breggen R, Suchiman HE, Lakenberg N, et al: **Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited.** *Aging Cell* 2011, **10:**686-698.

102. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, et al: **Genome-wide methylation profiles reveal quantitative views of human aging rates.** *Mol Cell* 2013, **49:**359-367.

103. Jaffe AE, Irizarry RA: **Accounting for cellular heterogeneity is critical in epigenome-wide association studies.** *Genome Biol* 2014, **15:**R31.

104. Teschendorff AE, West J, Beck S: **Age-associated epigenetic drift: implications, and a case of epigenetic thrift?** *Hum Mol Genet* 2013, **22:**R7-R15.

105. Katiyar SK, Matsui MS, Mukhtar H: **Kinetics of UV light-induced cyclobutane pyrimidine dimers in human skin in vivo: an immunohistochemical analysis of both epidermis and dermis.** *Photochem Photobiol* 2000, **72:**788-793.

106. McGrath JA, Uitto J: **Anatomy and Organization of Human Skin.** In *Rook's Textbook of Dermatology.* Wiley-Blackwell; 2010: 1-53

107. Fisher GJ, Wang ZQ, Datta SC, Varani J, Kang S, Voorhees JJ: **Pathophysiology of premature skin aging induced by ultraviolet light.** *N Engl J Med* 1997, **337:**1419-1428.

108. Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CP, van Dijk CM, Tollenaar RA, et al: **Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains.** *Nat Genet* 2012, **44:**40-46.

109. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al: **High density DNA methylation array with single CpG site resolution.** *Genomics* 2011, **98:**288-295.

110. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry R: **Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays.** *Biostatistics* 2013, **submitted**.

111. Hybbinette S, Bostrom M, Lindberg K: **Enzymatic dissociation of keratinocytes from human skin biopsies for in vitro cell propagation.** *Exp Dermatol* 1999, **8:**30-38.

112. Poumay Y, Roland IH, Leclercq-Smekens M, Leloup R: **Basal detachment of the epidermis using dispase: tissue spatial organization and fate of integrin alpha 6 beta 4 and hemidesmosomes.** *J Invest Dermatol* 1994, **102:**111-117.

113. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA: **Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies.** *Int J Epidemiol* 2012, **41:**200-209.

114. Hansen KD, Sabunciyan S, Langmead B, Nagy N, Curley R, Klein G, Klein E, Salamon D, Feinberg AP: **Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization.** *Genome Res* 2013.

115. Watson CT, Disanto G, Sandve GK, Breden F, Giovannoni G, Ramagopalan SV: **Age-associated hyper-methylated regions in the human brain overlap with bivalent chromatin domains.** *PLoS One* 2012, **7:**e43840.

116. Zhao XD, Han X, Chew JL, Liu J, Chiu KP, Choo A, Orlov YL, Sung WK, Shahab A, Kuznetsov VA, et al: **Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells.** *Cell Stem Cell* 2007, **1:**286-298.

117. Griffiths CE, Wang TS, Hamilton TA, Voorhees JJ, Ellis CN: **A photonumeric scale for the assessment of cutaneous photodamage.** *Arch Dermatol* 1992, **128:**347-351.

118. Helfrich YR, Yu L, Ofori A, Hamilton TA, Lambert J, King A, Voorhees JJ, Kang S: **Effect of smoking on aging of photoprotected skin: evidence gathered using a new photonumeric scale.** *Arch Dermatol* 2007, **143:**397-402.

119. Liu Y, Li X, Aryee MJ, Ekstrom TJ, Padyukov L, Klareskog L, Vandiver A, Moore AZ, Tanaka T, Ferrucci L, et al: **GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease.** *Am J Hum Genet* 2014, **94:**485-495.

120. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, Busche S, Yuan W, Nisbet J, Sekowska M, et al: **Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements.** *Am J Hum Genet* 2013, **93:**876-890.

121. Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris TB, Launer L, Gudnason V, Fallin MD: **Personalized epigenomic signatures that are stable over time and covary with body mass index.** *Sci Transl Med* 2010, **2:**49ra67.

122. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM: **Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking.** *Hum Mol Genet* 2013, **22:**843-851.

123. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102:**15545-15550.

124. Buonaccorsi JN, Prieto VG, Torres-Cabala C, Suster S, Plaza JA: **Diagnostic utility and comparative immunohistochemical analysis of MITF-1 and SOX10 to distinguish melanoma in situ and actinic keratosis: a clinicopathological and immunohistochemical study of 70 cases.** *Am J Dermatopathol* 2014, **36:**124-130.

125. Zilliox MJ, Irizarry RA: **A gene expression bar code for microarray data.** *Nat Methods* 2007, **4:**911-913.

126. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, et al: **Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer.** *Genome Res* 2010, **20:**440-446.

127. Fernandez AF, Bayon GF, Urdinguio RG, Torano EG, Garcia MG, Carella A, Petrus-Reurer S, Ferrero C, Martinez-Camblor P, Cubillo I, et al: **H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells.** *Genome Res* 2014.

128. Consortium EP: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489:**57-74.

129. Gilchrest BA, Blog FB, Szabo G: **Effects of aging and chronic sun exposure on melanocytes in human skin.** *J Invest Dermatol* 1979, **73:**141-143.

130. Leek JT, Storey JD: **Capturing heterogeneity in gene expression studies by surrogate variable analysis.** *PLoS Genet* 2007, **3:**1724-1735.

131. Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J: **Epigenome-wide association studies without the need for cell-type composition.** *Nat Methods* 2014, **11:**309-311.

132. Houseman EA, Molitor J, Marsit CJ: **Reference-free cell mixture adjustments in analysis of DNA methylation data.** *Bioinformatics* 2014, **30:**1431-1439.

133. Aryee M: **Comment on PMID: 24464286.** *PubMed Commons* 2014, **Internet**.

134. Blattler A, Yao L, Witt H, Guo Y, Nicolet CM, Berman BP, Farnham PJ: **Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes.** *Genome Biol* 2014, **15:**469.

135. Yan W, Zhang LL, Yan L, Zhang F, Yin NB, Lin HB, Huang CY, Wang L, Yu J, Wang DM, Zhao ZM: **Transcriptome analysis of skin photoaging in chinese females reveals the involvement of skin homeostasis and metabolic changes.** *PLoS One* 2013, **8:**e61946.

136. Takeuchi H, Runger TM: **Longwave UV light induces the aging-associated progerin.** *J Invest Dermatol* 2013, **133:**1857-1862.

137. Young SG, Yang SH, Davies BS, Jung HJ, Fong LG: **Targeting protein prenylation in progeria.** *Sci Transl Med* 2013, **5:**171ps173.

138. Ladd-Acosta C, Hansen KD, Briem E, Fallin MD, Kaufmann WE, Feinberg AP: **Common DNA methylation alterations in multiple brain regions in autism.** *Mol Psychiatry* 2013.

139. McCall MN, Bolstad BM, Irizarry RA: **Frozen robust multiarray analysis (fRMA).** *Biostatistics* 2010, **11:**242-253.

140. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3:**Article3.

141. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP: **GenePattern 2.0.** *Nat Genet* 2006, **38:**500-501.

142. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9:**357-359.

143. Dantas Machado AC, Zhou T, Rao S, Goel P, Rastogi C, Lazarovici A, Bussemaker HJ, Rohs R: **Evolving insights on how cytosine methylation affects protein-DNA binding.** *Brief Funct Genomics* 2015, **14:**61-73.

144. Maunakea AK, Chepelev I, Cui K, Zhao K: **Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition.** *Cell Res* 2013, **23:**1256-1269.

145. Timp W, Bravo HC, McDonald OG, Goggins M, Umbricht C, Zeiger M, Feinberg AP, Irizarry RA: **Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors.** *Genome Med* 2014, **6:**61.

146. Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE: **An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging.** *PLoS Genet* 2015, **11:**e1004996.

147. Kilpatrick RD, Rickabaugh T, Hultin LE, Hultin P, Hausner MA, Detels R, Phair J, Jamieson BD: **Homeostasis of the naive CD4+ T cell compartment during aging.** *J Immunol* 2008, **180:**1499-1507.

148. Kumar M, Lu Z, Takwi AA, Chen W, Callander NS, Ramos KS, Young KH, Li Y: **Negative regulation of the tumor suppressor p53 gene by microRNAs.** *Oncogene* 2011, **30:**843-853.

149. Bridges D, Ma JT, Park S, Inoki K, Weisman LS, Saltiel AR: **Phosphatidylinositol 3,5-bisphosphate plays a role in the activation and subcellular localization of mechanistic target of rapamycin 1.** *Mol Biol Cell* 2012, **23:**2955-2962.

150. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al: **Integrative analysis of 111 reference human epigenomes.** *Nature* 2015, **518:**317-330.

151. Pinello L, Xu J, Orkin SH, Yuan GC: **Analysis of chromatin-state plasticity identifies cell-type-specific regulators of H3K27me3 patterns.** *Proc Natl Acad Sci U S A* 2014, **111:**E344-353.

152. Bektas A, Zhang Y, Wood WH, 3rd, Becker KG, Madara K, Ferrucci L, Sen R: **Age-associated alterations in inducible gene transcription in human CD4+ T lymphocytes.** *Aging (Albany NY)* 2013, **5:**18-36.

153. Schraml BU, Hildner K, Ise W, Lee WL, Smith WA, Solomon B, Sahota G, Sim J, Mukasa R, Cemerski S, et al: **The AP-1 transcription factor Batf controls T(H)17 differentiation.** *Nature* 2009, **460:**405-409.

154. Schmitt V, Rink L, Uciechowski P: **The Th17/Treg balance is disturbed during aging.** *Exp Gerontol* 2013, **48:**1379-1386.

155. Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, Teschendorff AE, Butcher LM, Beck S, Muller M: **Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs.** *Age (Dordr)* 2014, **36:**9648.

156. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S, et al: **Methylation of ELOVL2 gene as a new epigenetic marker of age.** *Aging Cell* 2012, **11:**1132-1134.

157. Michie CA, McLean A, Alcock C, Beverley PC: **Lifespan of human lymphocyte subsets defined by CD45 isoforms.** *Nature* 1992, **360:**264-265.

158. Whitelaw DM: **Observations on human monocyte kinetics after pulse labeling.** *Cell Tissue Kinet* 1972, **5:**311-317.

159. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al: **DNA-binding factors shape the mouse methylome at distal regulatory regions.** *Nature* 2011, **480:**490-495.

160. Brandenberger C, Li N, Jackson-Humbles DN, Rockwell CE, Wagner JG, Harkema JR: **Enhanced allergic airway disease in old mice is associated with a Th17 response.** *Clin Exp Allergy* 2014, **44:**1282-1292.

161. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA: **Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.** *Bioinformatics* 2014, **30:**1363-1369.

162. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, et al: **Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis.** *Nat Biotechnol* 2013, **31:**142-147.

163. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Berdasco M, Fraga MF, O'Hanlon TP, Rider LG, et al: **Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus.** *Genome Res* 2010, **20:**170-179.

164. Gervin K, Vigeland MD, Mattingsdal M, Hammero M, Nygard H, Olsen AO, Brandt I, Harris JR, Undlien DE, Lyle R: **DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes.** *PLoS Genet* 2012, **8:**e1002454.

165. Singh AM, Chappell J, Trost R, Lin L, Wang T, Tang J, Matlock BK, Weller KP, Wu H, Zhao S, et al: **Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells.** *Stem Cell Reports* 2013, **1:**532-544.

166. Ly T, Ahmad Y, Shlien A, Soroka D, Mills A, Emanuele MJ, Stratton MR, Lamond AI: **A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells.** *Elife* 2014, **3:**e01630.

167. Takase K, Ohno S, Takeno M, Hama M, Kirino Y, Ihata A, Ideguchi H, Mochida Y, Tateishi U, Shizukuishi K, et al: **Simultaneous evaluation of long-lasting knee synovitis in patients undergoing arthroplasty by power Doppler ultrasonography and contrast-enhanced MRI in comparison with histopathology.** *Clin Exp Rheumatol* 2012, **30:**85-92.

168. Gold D, Mallick B, Coombes K: **Real-time gene expression: statistical challenges in design and inference.** *J Comput Biol* 2008, **15:**611-623.

169. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O: **Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells.** *Nat Biotechnol* 2015, **33:**155-160.

170. Hervouet E, Nadaradjane A, Gueguen M, Vallette FM, Cartron PF: **Kinetics of DNA methylation inheritance by the Dnmt1-including complexes during the cell cycle.** *Cell Div* 2012, **7:**5.

171. Brown SE, Fraga MF, Weaver IC, Berdasco M, Szyf M: **Variations in DNA methylation patterns during the cell cycle of HeLa cells.** *Epigenetics* 2007, **2:**54-65.

172. Desjobert C, El Mai M, Gerard-Hirne T, Guianvarc'h D, Carrier A, Pottier C, Arimondo PB, Riond J: **Combined analysis of DNA methylation and cell cycle in cancer cells.** *Epigenetics* 2015, **10:**82-91.

173. Polyak K, Hamilton SR, Vogelstein B, Kinzler KW: **Early alteration of cell-cycle-regulated gene expression in colorectal neoplasia.** *Am J Pathol* 1996, **149:**381-387.

174. Backus HH, Van Groeningen CJ, Vos W, Dukers DF, Bloemena E, Wouters D, Pinedo HM, Peters GJ: **Differential expression of cell cycle and apoptosis related proteins in colorectal mucosa, primary colon tumours, and liver metastases.** *J Clin Pathol* 2002, **55:**206-211.

175. Sinclair AJ, Palmero I, Peters G, Farrell PJ: **EBNA-2 and EBNA-LP cooperate to cause G0 to G1 transition during immortalization of resting human B lymphocytes by Epstein-Barr virus.** *EMBO J* 1994, **13:**3321-3328.

176. Eldridge SR, Goldsworthy SM: **Cell proliferation rates in common cancer target tissues of B6C3F1 mice and F344 rats: effects of age, gender, and choice of marker.** *Fundam Appl Toxicol* 1996, **32:**159-167.

177. Hayflick L: **The Limited in Vitro Lifetime of Human Diploid Cell Strains.** *Exp Cell Res* 1965, **37:**614-636.

178. Abdallah P, Luciano P, Runge KW, Lisby M, Geli V, Gilson E, Teixeira MT: **A two-step model for senescence triggered by a single critically short telomere.** *Nat Cell Biol* 2009, **11:**988-993.

179. Chandra T, Ewels PA, Schoenfelder S, Furlan-Magaril M, Wingett SW, Kirschner K, Thuret JY, Andrews S, Fraser P, Reik W: **Global reorganization of the nuclear landscape in senescent cells.** *Cell Rep* 2015, **10:**471-483.

180. Shah PP, Donahue G, Otte GL, Capell BC, Nelson DM, Cao K, Aggarwala V, Cruickshanks HA, Rai TS, McBryan T, et al: **Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape.** *Genes Dev* 2013, **27:**1787-1799.

181. Bork S, Pfister S, Witt H, Horn P, Korn B, Ho AD, Wagner W: **DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells.** *Aging Cell* 2010, **9:**54-63.

182. Cawthon RM: **Telomere measurement by quantitative PCR.** *Nucleic Acids Res* 2002, **30:**e47.

Curriculum Vitae

AMY RUTH VANDIVER
30 S. Regester Street
Baltimore, MD 21231
310-892-5819
amyruthvandiver@gmail.com

EDUCATION

Johns Hopkins University School of Medicine, Baltimore, MD August 2008- present
Expected completion of MD/PhD May 2017

Honors
Excellence in Medical Student Research Award 2010

Princeton University, Princeton, NJ September 2004- June 2008
A.B, Summa Cum Laude in Molecular Biology

Honors
Molecular Biology Senior Thesis Prize
Highest Honors in Molecular Biology
Sigma Xi Scientific Research Honor Society

RESEARCH

Department of Molecular Medicine, Johns Hopkins University School of Medicine, "Epigenetic change in human aging."

Advisor: Andrew Feinberg, M.D. August 2010- present

Department of Geriatrics, Johns Hopkins University School of Medicine. "DNA Methylation and regulation of gene expression in frailty."

Advisor: Jeremy Walston, M.D. August 2009- August 2010

Department of Molecular Biology, Princeton University. "Cell Fate Specification and Cell Coordination in Cephalic Furrow Formation in Drosophila."

Advisor: Eric Wieschaus, Ph.D. January 2007- May 2008

Department of Medicine, University of Illinois at Chicago School of Medicine. " Effects of FOXO1 on Gene Expression in Skeletal Muscle."

Advisor: Terry Unterman, M.D. June 2006-December 2006

PUBLICATIONS

Journal Articles

Vandiver AR, Irizarry RA, Hansen KD, Garza LA, Runarsson A, Li X, Chien AL, Wang TS, Leung SG, Kang S, Feinberg AP. Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome Biology.* 2015; 16:80. doi: 10.1186/s13059-015-0644-y.

Liu Y, Li X, Aryee MJ, Ekstrom TJ, Padyukov L, Klareskog L, **Vandiver A**, Moore AZ, Tanaka T, Ferrucci L, Fallin MD, Feinberg AP (2014) GeMes, Clusters of DNA Methylation under Genetic Control, Can Inform Genetic and Epigenetic Analysis of Disease. *Am J Hum Genet.* 2014 Apr 3;94(4):485-95.

PRESENTATIONS

Invited Seminars

Epigenetics and Dermatology. Grand rounds lecture for the Johns Hopkins Department of Dermatology. *Baltimore, MD.* April 23, 2014.

Scientific Meetings

Oral Presentations

Widespread genomic blocks of hypomethylation in nonmalignant epidermis related to age and sun exposure. Oral Presentation at the Society of Investigative Dermatology Annual Meeting. *Albuquerque, NM.* May 9, 2014.

Poster Presentations

Age and sun exposure contribute synergistically to epigenomic change in human epidermis. Poster presentation at the Centers of Excellence in Genomic Science (CEGS) Meeting. *Madison, WI.* October 15, 2013

Genome-wide epigenetic change is associated with epidermal aging and UV exposure. Poster presentation at the Centers of Excellence in Genomic Science (CEGS) Meeting. *Chapel Hill, NC.* October 15, 2012.

Genome-wide epigenetic change is associated with epidermal aging and UV exposure. Poster Presentation at the Epigenomics of Common Disease Meeting. *Baltimore, MD.* October 13, 2012.

Differential methylation in hematopoietic stem cell aging. Poster presentation at the Centers of Excellence in Genomic Sciences (CEGS) Meeting. *Boston, MA.* October 19, 2011.

DNA Methylation in the Frail Mouse Model. Poster presentation at the American Geriatrics Society meeting. *Orlando, FL.* May 12, 2010.

Effects of FOXO1 on Gene Expression in Skeletal Muscle. Poster presentation at American Diabetes Association Scientific Sessions. *Chicago, IL.* June 22, 2007.

MENTORING

Mark Sabbagh, first year graduate student, Summer 2012.
Paulina Davison, high school senior. Summer 2013.
Sophia Caldera, high school senior. Summer 2014.

WORK EXPERIENCE

Princeton University, Princeton, NJ: *Residential Advisor* August 2006- May 2008
- •Provide guidance and support to 30 freshmen and sophomore students
- •Coordinate two events per year to address diversity issues

UCLA School of Public Health, Los Angeles, CA: *Intern* Summer 2005
- •Compiled an electronic database of Monkeypox literature
- •Designed survey forms for use in field studies